

Managing and Sustaining A State Government Publications Program in California

**A Report on the Existing Situation and
Recommendations for Action**

prepared for the Metropolitan Cooperative Library System

by

**Judith Cobb
Gayle Palmer**

OCLC Digital Collection and Metadata Services Division

August 30, 2004

This publication was supported by the U.S. Institute of Museum and Library Services under the provisions of the Library Services and Technology Act, administered in California by the State Librarian.

The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Institute of Museum and Library Services or the California State Library, and no official endorsement by the U.S. Institute of Museum and Library Services or the California State Library should be inferred.

Table of Contents

About the Report	i
Executive Summary	iii
Section 1: Digital Government Information Architecture	1
Section 2: The Current Environment	7
Section 3: Components of a New State Repository Program	15
3.1 Legal Issues: Definitions, Authority and Compliance	15
3.2 Copyright Considerations	19
3.3 Data Capture	23
3.3a The Role of Content Producers (State Agencies) in Data Capture	25
3.3b Content Selection	29
3.3c Metadata	31
3.4 Data Management and Access	33
3.4a Access	33
3.4b Repository Architecture	34
3.4c Authenticity	38
3.4d File Formats	39
3.5 Digital Preservation	40
3.6 Program Administration	41
3.6a Roles and Responsibilities	44
Section 4: Moving Forward	47
Conclusion	51
Appendix 1: Final Report: California State Library Government Documents Conference	
Appendix 2: Additional Resources: States and Copyright	
Appendix 3: Bibliography	
Appendix 4: List of Interviews	
Appendix 5: List of Acronyms	

Managing and Sustaining a State Government Publications Program in California

About the Report

Scope and Purpose

This report articulates ongoing issues, recommendations, and appropriate actions to ensure that the California State Document Depository Library Program operates effectively within the digital environment. The report intends to broadly address larger issues regarding digital government and information in order to provide the context for a more specific discussion of how to effectively manage digital publications. The report focuses more on a fundamental and technology-independent organizational and legal framework for the depository program than on specific technology implementations.

Background

In early 2003, the California State Library began to address program and systemic problems that were preventing the California State Document Depository Library Program from performing its primary mission as stated in the Library Distribution Act (California Government Code section 14900), "...to make freely available to inhabitants all state publications by distribution to libraries throughout the state," and to maintain these publications "...readily accessible for use...without charge." The assumption in the law is that the public will have access to state publications through distribution of publications to depository libraries situated throughout the state. This assumption breaks down when the distribution system fails. For various reasons, primary among them the increasing use of Web publishing by state agencies over the past ten years, the distribution system has become ineffective as a means of preserving and providing access to state government information. In June 2003 the State Library partnered with the Metropolitan Cooperative Library System (MCLS) in Pasadena to submit a successful proposal for Library Services and Technology Act funding to conduct a project to address these issues.

In November 2003, MCLS issued a contract to the OCLC Digital Collection and Metadata Services (DCMS) division to conduct a research study on the California State Document Depository Library Program and report its findings. Working with an advisory committee, the researchers examined and evaluated the program in relation to similar programs in other states, nations and organizations, focusing on digital resources management. A second contract with DCMS provided for the planning and implementation of a complementary statewide conference amongst various stakeholders, including depository libraries and state government agencies, to discuss the issues surrounding preservation of and persistent access to digital state government information. This conference was held on March 24 and 25, 2004.

Audience

The primary audience for this report includes the management and staff of the California State Library; the California State Archives; the California Records Management Program; the Office of State Publishing; the state data centers; California depository libraries; the Governor's Office of Planning and Research; all agencies and programs of California government; the California legislature; and other programs and states that are going through a similar effort to plan and implement comprehensive electronic document programs.

Managing and Sustaining a State Government Publications Program in California

About the Investigators

Judith Cobb is the Preservation Projects Specialist for OCLC Digital Collections and Metadata Services. She consults on archival and technology issues; develops workshops and curriculum related to digitization and digital preservation; and manages Web content for the DCMS Education and Planning section. Judith has more than 10 years of experience working with government agencies, specializing in born digital resource management and preservation issues, including metadata, selection, Web site and Web document preservation.

Judith was formerly the Assistant State Archivist at the State Archives of Ohio. She has worked extensively on state and national digital preservation projects, including the Illinois State Library, the Minnesota State Archives, the San Diego Supercomputer Center, the Ohio Supercomputer Center, and The State Library of Ohio.

Gayle Palmer is Digital and Preservation Services Manager for OCLC Western in Lacey, Washington where she coordinates consulting, training and grant programs. She is active as a trainer and consultant for digital project management, metadata, and Web development for government information.

Gayle formerly served as Principal Librarian at the Washington State Library where she developed and coordinated Washington State's Government Information Locator Service, the Find-It! Washington and Find-It! consumer search engines and the Washington Statewide Digital Images Initiative.

Research Methods

The purpose of the research project was to conduct an environmental scan of the current practices and models in operation among other states, nations and programs; make appropriate recommendations as to the best practices and models, and raise issues that remain in question. Cobb and Palmer conducted a thorough search and review of research reports, ongoing research projects on this and related topics, reviewing best practices and model programs as well as professional literature from several communities. The investigators analyzed the results of the California State Library Government Documents Conference, held in Sacramento on March 24 – 25, 2004. The investigators also conducted interviews with 15 individuals and met with representatives of the depository libraries.

Acknowledgements

The investigators acknowledge the valuable assistance of the project's advisory committee in planning and preparing this research report. The members of the committee were: David Cismowski, Janet Coles, Debbie Newton, Kristine Ogilvie, and Deborah Weber of the California State Library; Lucy Barber of the California State Archives; Guy Blair of the California Office of State Publishing; Ben Amata of the California State University, Sacramento; and Linda Kennedy of the University of California, Davis. The investigators particularly thank Janet Coles for her vision and thoughtful leadership during the research project.

Executive Summary

Government information serves many purposes. It is used for the development of policy, for the evaluation of government, for education, for determining political boundaries and for the allocation of resources. It is also used to serve the democratic process. Government information is in demand from all sectors, from legislators and policy-makers, from research and development entities, from the commercial sector, from grassroots and nonprofit agencies and from the public. Its value is inestimable, and the costs of not having it are incalculable.

Recognizing this, the California State Legislature established the California State Document Depository Library Program in 1945. Managed by the California State Library since 1959, the program's mission is to preserve all California state government publications and make them permanently and publicly available through a depository library system. Now that most state government publications are available only through the World Wide Web, the program fails to fulfill its mission because there are few mechanisms in place to preserve those digital publications and provide access to them over time, or even to notify librarians and the public about their existence. The resultant loss of state government information is untold.

A new repository program that effectively preserves, manages, and provides persistent public access to digital state government published information should be created for California. Such a program should function within a larger state government information infrastructure. "Information infrastructure" refers to a comprehensive and integrated implementation of law, policy, programs and workflows to preserve vital government information; it encompasses records management, technology implementation, system design, information creation, and long-term preservation. This information infrastructure should be an integral part, along with a technological infrastructure, of an overall digital government information architecture. In the push for digital capability over the past 15 years, governments have often ignored the components of an information infrastructure in implementing information technology, and this is also true for California.

Defining an overall information infrastructure for California state government is beyond the scope of this report. The new repository program, however, should perform as part of such an infrastructure. Ideally the new program would be created in conjunction with infrastructure development. Because of this the new program must be planned and implemented using tested and accepted policies and standards as much as possible, so that it may serve as a model for development of the larger information infrastructure.

Core considerations for building a new repository program are as follows:

- 1) *Partnership*. The effort to build a new program will be complex and multifaceted and must be accomplished through a partnership of stakeholders, as no single organization has the expertise or resources to accomplish it on its own.
- 2) *Legislative change*. New legislation will be necessary to redefine the notion of state government published information in the digital environment and to create necessary incentives for compliance.

Managing and Sustaining a State Government Publications Program in California

- 3) *Copyright considerations.* Authoritative legal opinions and perhaps legislative changes concerning copyright and permissions will need to be obtained to ensure that permanent public access and preservation of digital state government information are legally acceptable.
- 4) *Data capture and management.* The program should incorporate multiple models for acquisition and management of digital state government published information based on an environment of collaboration with content producers (state agencies). These models might include active distribution from the agency, active notification from the agency for manual harvesting, and selected or bulk automatic harvesting, and should include training and outreach to agencies in the area of program guidelines and standards, including standards for descriptive, structural and technical metadata. Much thought should be put into the development of guidelines and standards, including selection and authenticity of content, methods of storage and preservation, and portals for access. End users as well as content producers should be involved in this development.
- 5) *Digital preservation.* To sustain digital government information resources over time without losing significant informational and contextual properties, an Open Archival Information System (OAIS) compliant digital archive should be implemented as a component of the new program.
- 6) *Program administration and governance.* Key stakeholders should meet and begin defining roles, responsibilities and relationships concerning administration of the new program. Stakeholders would commit to establishing the new program, plan for collaboration by establishing committees and working groups, and establish memorandums of understanding identifying specific roles and responsibilities.
- 7) *Funding and sustainability.* A plan for funding this new program should be drawn up. Sources of funding might include: grant funding for technology test bed projects; participation as partners in projects with other states and the federal Government Printing Office; consideration of a fee-for-service cost model; seeking operational funding from stakeholder organizations; and development of a long-range funding request from stakeholder agencies to the California legislature. The fiscal consequences of failure to act should be documented where possible.

Developing a new repository program will not be simple. It will require coordinated action by the California State Library working with key partners including the California State Archives, the Department of General Services, the state data centers, state agencies, California state depository libraries and other stakeholders such as the California Digital Library and the San Diego Supercomputer Center. California already has organizations with the knowledge and tools to do the job; it must work to bring them together to assure that California digital state government information is permanently accessible and usable. The disappearance of state government information has present and future implications for every Californian. Preservation of, and permanent public access to, this information is imperative; the state's historical, cultural, and intellectual record is at stake.

Section 1 Digital Government Information Architecture: Establishing an Information Infrastructure

Digital government has transformed the ways in which California government agencies interact with citizens and make government information available to the public. At the heart of digital government initiatives is a commitment to broad access to information and services as well as a belief that technology can save the government money.¹ In order to create an environment in which digital government can flourish, the state of California has invested significantly in the technological infrastructure that is the backbone of its electronic environment. The current technological infrastructure supports a variety of digital government services, from the state portal to e-mail for state employees, and requires approximately two billion dollars each year from the state budget.² These systems, consisting of hardware, software and applications, are networked together to make digital government work more effectively and be more accessible. They also hold a vast resource that is often taken for granted – information.

Worldwide, in the rush to create Web sites and services, to “go digital,” governments have focused on technological infrastructure and ignored the information infrastructure. What is the “information infrastructure?” It is policy, procedures, programs, and technology all working together to ensure that the government’s information is reliable, managed effectively, and accessible. Digital information, whether stored on large back-end systems or Web portals, represents a vast investment of the resources of government. These are resources that should not be mismanaged.

The value of this information is immense, not just to government but also to the corporate sector. In an article on the topic, George J. Symons wrote, “Today, the flow of information is a critical factor in enabling an organization to meet its core business objectives. Strategic information flow is helping to drive competitive advantage in terms of reduced cycle times, lower operational costs and expanded revenue opportunities. Managing information assets now takes on greater importance than ever before.”³

Managing information, whether analog or digital, implies a commitment to understanding the value of the information and maintaining it for as long as is appropriate. For example, data collected from a citizen who goes online to order vital records or make appointments may need to be retained in the system for only a short time, while legislative records, created and stored digitally, need to be maintained permanently. Environmental reports that reflect vital information and an investment of government research and resources also need to be maintained for the long term. If these important pieces of government information are mismanaged, vital information will be lost or have to be recreated at taxpayers’ expense.

¹DeMaio, Carl, et al. *Citizens Budget, 2003 – 2005*. Los Angeles, CA: Reason Public Policy Institute and the Performance Institute, 2003. Available from <http://www.performanceweb.org/pi/research/california/sectionone-page3.htm>; Internet; accessed 30 August 2004. The executive summary to this report states, in point 14 that information technology can be used to “create cost efficiencies, reduce duplication and overhead, and improve cross-program coordination.”

² Ibid., p. 47.

³ George J. Symons, “Assessing the Value of Content: Information Management,” *DM Review Magazine*, March 2004; available from http://www.dmreview.com/article_sub.cfm?articleId=8175; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

In an attempt to gain more understanding of this issue, the U.S. National Commission on Libraries and Information Science (NCLIS) published a report, *Comprehensive Assessment of Public Information Dissemination*. The report examined assumptions about the value of government information. The authors found that:

People need access to government information to learn about what their government is doing. They need access to a vast range of information to make their lives better. They need to know of potentially harmful situations and events such as air pollution conditions, dangerous storms, or faulty tires. Researchers, scholars, and product developers need ready and timely access to a wide variety of scientific, technical, and business-related information and data. **The cost to citizens of not having information readily available cannot be calculated.**⁴

The report goes on to recommend legislation that would "formally recognize and affirm the concept that public information is a strategic national resource."⁵

The *Citizens' Budget, 2003–2005*, a private California budget plan issued by the Reason Public Policy Institute and the Performance Institute, recognizes the importance of creating an information infrastructure. In the section in Part 4 titled, "Consolidating Administrative Services into a 'Shared Services' Operation," it advocates the implementation of shared e-records management within California government. Such management, it says, has shown significant results in increased efficiency and cost savings in the federal government. Shared e-records management, the report states, "provides policy guidance to help agencies to better manage their electronic records, so that records information can be effectively used to support timely and effective decision-making, enhance service delivery, and ensure accountability. This...will improve the government's ability to ensure the integrity of electronic records and related information that agencies require to meet their legal and internal business needs."⁶

Ensuring governmental accountability is a right of each citizen. As California Government Code Section 54950 states:

The people of this State do not yield their sovereignty to the agencies which serve them. The people, in delegating authority, do not give their public servants the right to decide what is good for the people to know and what is not good for them to know. The people insist on remaining informed so that they may retain control over the instruments they have created.

⁴ U.S. National Commission on Libraries and Information Science, "Appendix 25, Panel 3. Final Report on Citizen, Business, Lower Levels of Government, Library, and Other Needs for Public Information Products and Services." In *A Comprehensive Assessment of Public Information Dissemination* (Washington, DC: U.S. National Commission on Libraries and Information Science, 2001), vol. 3, p. 124-135. Available at <http://www.nclis.gov/govt/assess/assess.appen25.pdf>; Internet; accessed 30 August 2004.

⁵ Ibid.

⁶ Carl DeMaio, et al., op cit., part 4.

Managing and Sustaining a State Government Publications Program in California

Making government information and services available via the Web certainly helps fulfill this fundamental mandate. The 2002 Pew Internet & American Life Project report, *Rise of the e-citizen: how people use government agencies Web sites*, states that overall, 60% of government Web site users say such sites had improved their interactions with at least one level of government.⁷ The section on California in a later report by the same organization, *Internet use by region in the United States*, shows that Web based services and Internet access to digital government information are quickly becoming essential to the daily activities of California inhabitants. Users in Los Angeles and San Francisco indicated that the ca.gov and state.ca.us domains were among their favorite Web sites.⁸

Building and supporting the technological infrastructure for California is important and has, rightly, been a fundamental task over the last ten years as technology has changed rapidly. However, the state must now focus on managing the information and records that its technological infrastructure creates and stores. If the information is mismanaged or left unmanaged over time, if access disappears, the fundamental mandate stated above cannot be met.

In summary, an effective digital government information architecture must include both a technological infrastructure (consisting of networks, hardware, software, etc.) and an information infrastructure (consisting of laws, administrative rules, standards, etc.). This information infrastructure forms the backbone for digital government and should be relatively static and independent of specific technology. The technological infrastructure will perpetually change, but should always meet the requirements of the information infrastructure.

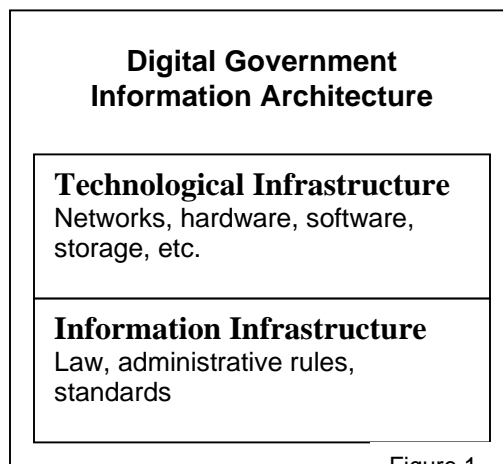


Figure 1

What Kinds of Information?

“E-records management,” as referred to in the Citizens’ Budget, encompasses records created in digital formats. However, in considering an information infrastructure for the state of California, we must not lose sight of the fact that, even today, not all information is digital. State agencies continue to create and manage paper-based records and publications. The information infrastructure must take into account all kinds of government information, not only digital formats. Further, in the digital environment, information management is so interconnected with technology that implementation becomes a collaboration between data creators, holders, and information professionals. The term “information infrastructure” implies a more comprehensive

⁷ Elena Larsen and Lee Rainie, *The rise of the e-citizen: how people use government agencies web sites* (Washington, DC: Pew Internet & American Life Project, 2002); available at http://www.pewinternet.org/pdfs/PIP_Govt_Website_Rpt.pdf Internet; accessed 30 August 2004. Overall, 60% of government Web site users say such sites had improved their interactions with at least one level of government. Half of government Web site users (49%) say the Internet has improved the way they interact with the federal government; 44% say it has improved the way they interact with their state government; and 30% say the Internet has improved the way they interact with local government.

⁸ Tom Spooner, *Internet use by region in the United States* (Washington, DC: Pew Internet & American Life Project, 2003), p. 93; available at http://www.pewinternet.org/pdfs/PIP_Regional_Report_Aug_2003.pdf; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

and integrated implementation of several related functions, including records management, digital publishing, technology implementation, system design, information creation, and the long-term preservation of vital information.

Defining an overall information infrastructure for California state government is beyond the scope of this report, which focuses on the development of a “new” depository library program that can operate effectively in the digital age. This new program, however, should be designed to function as part of such an infrastructure. In determining the components and functional requirements of the new program, then, the eventual needs of the larger information infrastructure in terms of policy and standards should be kept in mind.

California Government Publications and the Depository Library Program

Librarians in the Government Publications Section of the California State Library estimate that “...80% to 90% of all California state publications are now issued on the Web. For many of these there is no paper counterpart.”⁹ The cost savings and ease implicit in Web publishing, while providing real immediate benefits, create a number of very real dangers that directly impact citizen and government access to information over time. Digital publishing disperses central coordination and retention of documents and contributes to the breakdown of traditional print repositories. Loss of access to digital publications through lack of a comprehensive, coordinated program of action to identify and retain important digital output is a significant danger.

At the federal government level, the need for access to historical information is documented by the experience of the National Technical Information Service (NTIS), an agency that must supply information to the public on a cost-recovery basis. In the previously mentioned NCLIS report, the authors observe, “...statistics show that nearly two-thirds of all requests for NTIS materials are for materials that are over three years old, and one-third of the requests are for materials that are over 10 years old.”¹⁰

By their very nature digitally published documents are dynamic, volatile, and uncontrolled. California state agencies are not consistently planning for how users will access information if it is removed from an active Web site. Access to Web information can also be affected by “link rot.” This term refers to information that has become inaccessible because of an invalid link to a Web page, deletion or removal of Web sites, and/or loss of access to information previously published on Web pages. Janet Coles, California Documents Librarian at the California State Library, recently stated, “Documents and other resources are often removed as more current projects demand space and attention on agency Web pages.”¹¹

⁹ California State Library, “Enhancing Access to Electronic Resources Through the CSL Catalog: A Vision Statement,” 2003 internal publication.

¹⁰ F. Woody Horton, “Appendix 15. Some Important Information [on] Age Paradigms Shifts and Their Associated Myths and Realities.” In *A Comprehensive Assessment of Public Information Dissemination*, op cit., vol.3, p.14-24; available at <http://www.nclis.gov/govt/assess/assess.appen15.pdf>; Internet; accessed 30 August 2004.

¹¹ Janet Coles, California Documents Librarian, California State Library, Speech, 6th Annual State-GILS Conference, Raleigh, NC, 31 March 2004.

Managing and Sustaining a State Government Publications Program in California

Not only are Web publications endangered, but under the current California depository program only a small percentage of *printed* state publications are being identified and received in depository libraries. Print publications intended to come through the depository library system (as outlined in the Library Distribution Act) are not reaching the public. Agency awareness of the need for action relating to preservation and access to both digital and print resources is dangerously low.

Not only is ongoing public access to state government information hindered, but the state of California is also losing its valuable information resources. The information contained in the state's records, Web sites and publications takes enormous public resources to create; left unmanaged, the information is lost and the resources squandered. Lack of access to necessary information from both current and historical publications can significantly impact California government policy decisions and actions. There is a clear need to establish a permanent repository for public access to California government information resources, within the framework of a state information infrastructure.

During the California State Library Government Documents Conference, held in March 2004, Dean Misczynski, director of the California Research Bureau, noted that major government policy is frequently developed over several years, with legislators and policy staff making use of current and historical documents, issue papers and legislative testimony as they craft major policy initiatives. He noted that librarians must be "active participants" in assuring that the necessary documents remain accessible to the legislative process.¹²

It is important to recognize that information managers and producers are key stakeholders and must actively participate in the development and maintenance of the information infrastructure. This will include efforts assuming responsibility for obtaining necessary resources within California government.

To effectively provide a new focus and mission for a revitalized and ongoing government publications program, the project advisory committee has formulated a new vision statement:

To ensure that California government publications in all formats are usable, accessible, managed, and preserved via a collaborative, sustainable, comprehensive, and systematic program led by the California State Library for the inhabitants of California.

The first step toward achieving this vision is for the State Library and other key stakeholders to design a new repository program that ensures this mission can be carried out within the larger information infrastructure for California government. Achieving this goal will require a plan of action that includes:

¹² Dean Misczynski, Director, California Research Bureau., Speech, California State Library Government Documents Conference, 24 March 2004.

Managing and Sustaining a State Government Publications Program in California

Legislation/regulation:

- Strengthening the regulatory and legal mandate to manage the official and permanent repositories of government information by sponsoring changes to the State Administrative Manual and Library Distribution Act.
- Establishing guidelines and practices for information creators and managers.

Collaborative leadership:

- Designing a collaborative leadership model for California and identifying stakeholders that can share the responsibility and resources for both planning and implementation.
- Examining current roles and, where appropriate, establishing new roles and program responsibilities for the California state government agencies responsible for access and preservation of state government information, to support the new depository program.
- Identifying and involving other potential stakeholders and partners

Coordinated program development:

- Re-establishing and strengthening the vital network of communications with the information creators and managers.
- Creating a collaborative test-bed for technical issues and seeking partners to contribute to solutions.
- Seeking sustainable and ongoing technical and financial resources to support a comprehensive, coordinated program for the depository library component of the digital government information architecture.

Section 2 The Current Environment

This section outlines the context in which the California State Document Depository Library Program exists today, including the current legal environment, the current program, and the related organizations within the state that have that have potential stakeholder and/or partnership roles.

Size and Scope of the California Government

The California Legislative Analyst's Office states in a 2002 report, "California is the world's fifth largest economy."¹³ This is an indication of the size and power of California, and also of a correspondingly large and diverse governmental organization. The most current California Roster (2003 edition) describes 178 official listings for state departments, boards, and commissions.

In the online environment, the California state portal provides Web links to 961 state programs, agencies, boards and commissions. At this time no comprehensive scoping of the ca.gov domain has been undertaken. In their report, *Web-Based Government Information: Evaluating Solutions for Capture, Curation and Preservation*, the California Digital Library (CDL) described the government Web as "large, volatile, and opaque."¹⁴ In researching this report, CDL conducted test crawls of government Web sites, including some California sites. The test indicated that four file types, HTML, PDF, GIF and JPEG, comprised 95% of the content, although 335 file types were recognized.¹⁵

The California economy is currently in recovery, but government revenues have not kept up with demand for services. A budget deficit of more than 17 billion dollars is causing reductions in allocations to all agencies. The Schwarzenegger administration has placed a high priority on government reform and recently acted to consolidate and redefine responsibility for key technical initiatives as recommended by the California Performance Review (CPR), through consolidation of the Teale Data Center and the Health and Human Services Data Center.¹⁶

During the March 2004 conference, participants discussed problems with the current depository program. The general opinion was that lack of specific and focused language in existing legislation and regulations contributes to the ineffectiveness of the distribution and retention programs for California state publications. However, the participants found that other factors, such as a lack of coordinated leadership and a comprehensive plan of action, have the most immediate impact in the current environment.¹⁷

¹³ *CAL Facts: California's Economy and Budget in Perspective* (Sacramento, CA: Legislative Analyst's Office, December 2002), p. 1; available at http://www.lao.ca.gov/2002/cal_facts/cal_facts_2002.pdf; Internet; accessed 30 August 2004.

¹⁴ Patricia Cruse, et al., *Web-Based Government Information: Evaluating Solutions for Capture, Curation, and Preservation* (Oakland, CA: California Digital Library, November 2003), p. 13; available at <http://www.cdlib.org/inside/projects/preservation/govinfo/>; Internet; accessed 30 August 2004.

¹⁵ *Ibid.*, p. 11-12.

¹⁶ California Governor (Schwarzenegger), *Executive order S-13-04 by the Governor of the State of California* (Sacramento, CA: Office of the Governor, Aug. 24, 2004); available from <http://www.governor.ca.gov/> [click on "Press Room on left menu bar, then on "Executive Orders"]; Internet; accessed 30 August 2004.

¹⁷ See Appendix 1 of this report, "Final Report, California State Library Government Documents Conference."

Managing and Sustaining a State Government Publications Program in California

Legal Context

The California State Document Depository Library Program was established by law in 1945 and has been managed by the California State Library since 1959.¹⁸ It exists within a larger legal context that outlines how information resources are to be managed, preserved and made accessible to the public. This section will review the legal basis for the depository program, as well as the larger legal context of information management within the state.

Legislation and the State Depository Program

The Library Distribution Act (LDA) of 1945 established a system of state document depository libraries, and mandated that state agencies distribute their public documents to these libraries. The LDA is codified as California Government Code sections 14900-14912, and outlines state policy for public access to government documents as follows:

It is the policy of the State of California to make freely available to its inhabitants all state publications by distribution to libraries throughout the state, subject to the assumption by such libraries of the responsibilities of keeping such documents readily accessible for use, and of rendering assistance in their use to qualified patrons without charge.¹⁹

Section 14902 of the LDA broadly defines “state publication” to include nearly all documents produced by any state government agency that are not specifically meant for intra-agency use. It also defines “print” to include “all forms of duplicating other than by the use of carbon paper.”²⁰ Specific exclusions are noted in the law, such as interoffice memos and forms. From the outset the publications of the University of California were excluded from the program.

The State Administrative Manual (SAM) is a document produced by the California Department of General Services and the California Department of Finance to provide uniform guidance to state employees on completing their assignments. The SAM chapter dealing with agency responsibilities under the LDA refers to “state publications” as those of “lasting interest and use to the public.”²¹ This portion of the SAM has not been substantially updated since 1993.

Distribution of State Publications

The LDA requires agencies to distribute their publications to depository libraries, but does not prescribe a particular method. Agencies have the option to either send their publications directly to the libraries, or through the Office of State Publishing (OSP). If OSP is chosen as the distributor, the law establishes specific rules by which OSP must complete this distribution. In either case the decision to send a publication to depositories is left to the discretion of the creating agency.

¹⁸ *California state publications: manual for acquisition, processing, use* (Sacramento: California State Library, Government Publications Section, 1961), p. 3.

¹⁹ *California Government Code*, section 14900.

²⁰ *California Government Code*, section 14902.

²¹ *California State Administrative Manual. Chapter 3100, Publications and Documents* (Sacramento, CA: California Dept. of General Services), section 3121; available at <http://sam.dgs.ca.gov/TOC/3100/3121.htm>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

California government agencies have little incentive to comply with the present LDA, nor do they have any incentive to communicate or cooperate with the OSP or the State Library about their analog or digital publishing practices. As the law and regulations are currently written, there is neither benefit for agencies that comply nor sanction for non-compliance. Apathy and disinterest on the part of agencies in meeting obligations of the LDA are not new, nor are they the result of new publishing options in the digital era. Lack of compliance with the LDA was noted by state documents librarians as early as 1973.²²

The Depository Program and the Broader Legal Environment for Information Management and Access Within California

The design of the original depository program model was based on the notion that residents of California should have ready access to state publications from many library locations. Over the decades, the system outlined in the LDA for designating depository libraries to receive the publications was expanded. Much attention was paid to which libraries could be depositories and to establishing guidelines for “full” [receiving all publications] and “selective” [receiving selected publications] depository status. Regulations were published to describe the responsibilities of the depository libraries. At present, there are 134 depository libraries: 16 full depositories, including the California State Library, and 118 selective depositories.

The existing LDA language and SAM regulations have proven ineffective to meet the original legislative intent to assure free access for citizens to state publications distributed to California libraries. In a recent speech describing the current California state documents depository program, Janet Coles stated:

The California depository program has never really functioned well, and distribution problems have worsened over the past fifteen years and especially over the past five years. With the advent of Web delivery of government information, the program faces a severe crisis, as more and more agencies consider their obligation to provide information to the public satisfied by posting this information on their Web pages.²³

Currently, depository libraries do not receive California state government publications, whether published in tangible formats or in digital format on the Web. State agencies utilize the OSP with decreasing frequency; thus, what was intended to be a major distribution hub is no longer functioning as such.

California state agencies apart from the State Library that have broad mandates and responsibilities regarding access, preservation and management of state government records and publications include the California Department of General Services and the California Secretary of State. The Department of General Services manages the Office of State Publishing and also the California Records Management Program. The Records Management Program has two main

²² Linda Kennedy and Rosita Lo Russo, “The Crisis in the California State Documents Distribution System,” 1973 unpublished paper.

²³ Janet Coles, *op cit.*

Managing and Sustaining a State Government Publications Program in California

elements: California Records and Information Management (CalRIM) and the State Records Center. CalRIM establishes guidelines and provides training for state agencies to manage their records programs, including the management of electronic records and use of imaging technologies; the State Records Center offers low cost storage of vital records and semi-active and inactive records. The State Records Center is closed at the present time. Under the Secretary of State, the California State Archives collects, catalogs, preserves, and provides access to the historic records of state government and some local governments.

Defining a new program for access, preservation and management of digital California government publications will require that all of these agencies and programs work together with the State Library to define roles and responsibilities. They will also need to work with content producers and publishers in state agencies, and those who plan and manage information technology within state government, such as the recently formed Information Technology Council and the state data centers.

Laws Promoting Open Access and Transparency for California Government

In her presentation at the March 2004 conference, Joan Allen-Hart of the San Diego Public Law Library pointed out that the California legislature has consistently enacted legislation to support the right of citizens to free and open public access to government information.²⁴ This is one of the most positive aspects of the current environment in California. This recognition of the necessity of government transparency is reflected in the California Public Records Act²⁵; the Legislative Open Records Act²⁶; the State Records Management Act²⁷ and the Online Disclosure Act,²⁸ among others. Each of these laws affirms the concept of an informed citizenry and the public's right to access government information. While the history of California government policy favors the public's right to know, the current laws fall short in that they do not specifically address permanent public access to, and preservation of, California digital government information.

Current State Documents Program: Status Report

In 2002, the Office of State Publishing distributed 204 items to depository libraries, which it estimates was less than 5% of agency publications for that year.²⁹ In a similar estimate, librarians in the Government Publications Section of the California State Library estimate that "...80% to 90% of all California state publications are now issued on the Web. For many of these there is no paper counterpart."³⁰ Further, Janet Coles reports that over 95% of the print documents received have electronic counterparts on the World Wide Web.³¹ Clearly, the depository library program

²⁴ Joan Allen-Hart, San Diego Public Law Library, Speech, California State Library Government Documents Conference, 24 March 2004.

²⁵ *California Government Code*, Section 6250-6270.

²⁶ *California Government Code*, Section 9070-9080.

²⁷ *California Government Code*, Section 14740-14741.1.

²⁸ *California Government Code*, Section 84600-84612.

²⁹ Guy Blair, Administration Manager, California Office of State Publishing, Interview, 5 December 2003.

³⁰ California State Library, "Enhancing Access to Electronic Resources Through the CSL Catalog: A Vision Statement," 2003 internal publication.

³¹ *Ibid.*

Managing and Sustaining a State Government Publications Program in California

is not effectively meeting its mandate to “make freely available to [California] inhabitants all state publications” regardless of format.

California State Library, Government Publications Section

In addition to managing the California State Document Depository Program, the Government Publications Section (GPS) is the only regional depository of federal government information in California and is also one of five patent and trademark depository libraries in the state. The GPS staff is down from 17 FTEs in 2002 to 9.5 FTEs at present. At present only two of these positions are allocated to the California state documents collection and the state depository program. Despite limited staff, they do their best to make state agencies aware of the program and to encourage them to comply with the provisions of the Library Distribution Act. The State Library has undertaken several activities to address the current digital publishing environment.

In September 2003, the library began capturing and storing selected electronic publications by storing them on a network attached storage (NAS) server. Janet Coles reports that the NAS server has been used primarily to store publications from the governor’s site (because of the change in incumbents in November 2003) and publications from agencies that are closing due to legislative mandates or budget cuts. Materials from the administrations of Pete Wilson and Gray Davis have been collected and thousands of agency documents have been captured (about 600 MB total server space).³²

The Government Publications Section staff leads the effort to create metadata standards and implement metadata projects for California state agencies. The GPS staff also participates in the University of California shared cataloging program to establish composite records for government publications and add uniform resource identifiers for Web based resources to these records.

A key effort of the State Library is to provide bibliographic access to state government publications, both print and digital. The State Library provides original cataloging for California state documents and publishes a monthly acquisitions list on its Web site at <http://www.library.ca.gov/CSP/index.cfm>. This list of records, containing information on personal and agency authorship as well as subject content, provides one of the few records for posterity of the existence of a state publication. It serves as an alert to all libraries about titles they may want to acquire or provide access to for local users.

The State Library also maintains responsibility for content organization and management of the California state portal, <http://www.ca.gov/state/portal>. Although not directly related to the government documents area, this initiative has made the State Library a more visible resource for state government.

During the March 2004 conference, the State Library was acknowledged to be the primary organization that has both the knowledge and the responsibility to help lead an initiative to develop a new program to manage digital state government publications. Conference participants also noted frequently that the State Library could not accomplish so large a task on its own.

³² Janet Coles, Written comments, November 2003.

Managing and Sustaining a State Government Publications Program in California

There was widespread recognition that a collaborative model of leadership to include state agencies as content producers, the State Archives and the state data centers would be required.³³

Office of State Publishing

Operating under the California Department of General Services, the Office of State Publishing (OSP) offers printing solutions and services to California state agencies on a fee-for-service business model. Under the terms of the Library Distribution Act, the Office of State Publishing distributes agency-designated public documents to depository libraries. The OSP has maintained a close working relationship with the State Library as both have made a commitment to provide services to depositories. The OSP has also been significantly impacted by the advent of direct digital publishing by the agencies. Agencies are reducing the amount of print copies of publications and relying on direct access to information via the Web. Even important documents, such as the register of current contracts and the official state budget, are now produced only in digitally accessible formats.³⁴ The OSP archives electronic copies of the publications they handle in order to maintain production files for recurring publications.

Discovering New Partners

In order to effectively understand the existing environment and begin to consider potential solutions, one must be aware of the other organizations, specifically within California, that are working on related issues. As already stated, building a new depository program for California will require input, leadership, and responsibility from several stakeholders. Several of the primary stakeholders have already been mentioned: the State Library, State Archives, California Records and Information Management (CalRIM), and the Office of State Publishing. However, there are several other organizations that must also be considered as vital partners in this initiative.

State Data Centers

The two main data centers, the California Health and Human Services Agency Data Center (HHSDC)³⁵ and the Teale Data Center³⁶, act as multi-purpose centers that provide data processing services, network solutions, training, and information technology project management. As mentioned above, their consolidation was recently mandated by executive order. The Legislative Data Center provides similar services in support of the legislature and staff.

The technical infrastructure (software, connectivity hardware, etc), services, and expertise of the data centers are vital to creating and managing both the new program and the larger digital government architecture for the state. Representatives from the centers have already expressed interest in participating in an initiative to create a new depository program for digital government

³³ See Appendix 1 of this report, "Final Report, California State Library Government Documents Conference."

³⁴ Guy Blair, op cit.

³⁵ California Health and Human Services Agency Data Center, *California Health and Human Services Agency Data Center* [home page]; available at <http://www.hwdc.cahwnet.gov/about.asp>; Internet; accessed 30 August 2004.

³⁶ Stephen P. Teale Data Center (Calif.). *Teale: Stephen P. Teale Data Center* [home page]; available at <http://www.teale.ca.gov/>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

information. While the data centers will be key players in the development of a new depository program, the specific and appropriate responsibilities of the data centers will need to be determined by the stakeholders. It seems likely that at a minimum, the data centers could store and manage data for the new program, and possibly have a role in procurement.

The recently issued California Performance Review calls for two actions significant to this initiative regarding the state data centers: the aforementioned consolidation of the Teale and Health and Human Services Data Centers and a recommendation that the two centers immediately begin expanding their server hosting and management services and aggressively marketing them to state agencies. This consolidation of data and services lends itself to the possibility of new services, such as metadata generation and data storage for digital documents, which could be worked into the development of the new repository program.

The California Performance Review report also recognizes the relationship between the data centers and the oversight role played by the Department of Finance and the Chief Information Officer, J. Clark Kelso. The ability of the information management agencies (the State Library, State Archives and CalRIM) to select and purchase software and equipment to support the goals of the new depository program is significantly impacted by the oversight and regulations imposed by the Department of Finance. In the current California environment progress toward development of a new depository program is inherently tied to the outcome of the recommendations of the California Performance Review.

California Digital Library

The California Digital Library (CDL) is a program of the University of California.³⁷ The CDL is actively undertaking an effort to develop a suite of tools for Web archiving that would be available within and without the UC system for operation of automatic capture and Web archiving activities. As part of a Mellon Foundation-funded project, the CDL hopes to develop and test tool sets that can be widely used by libraries and organizations in the University of California (UC) system to implement digital preservation programs.³⁸ The goals and activities of this project are specifically outlined in a project report from initiative to research Web based government information.³⁹ Additionally, the CDL shared cataloging program assists in providing bibliographic records for digital publications.

San Diego Supercomputer Center

The San Diego Supercomputer Center (SDSC), a research unit of the University of California, San Diego, has worked for several years with the National Archives and Records Administration (NARA) and several state archival programs, including Ohio, Michigan, Minnesota, Kentucky,

³⁷ California Digital Library. *California Digital Library: Explore Digital Information from the University of California* [Web site]; available at <http://www.californiadigitallibrary.org/> Internet; accessed 30 August 2004. "The CDL is a collaborative effort of the ten University of California campuses. Drawing upon expertise from across the UC system, it selects, designs, builds, and manages systems for the use and preservation of high-quality digital content. The California Digital Library also works together with California's libraries, archives, museums, and other memory organizations to provide access to the cultural and historical resources of California."

³⁸ Patricia Cruse, Director, Digital Preservation Program, California Digital Library, Interview, May 2004.

³⁹ Patricia Cruse, et al., op cit.

Managing and Sustaining a State Government Publications Program in California

and Kansas, to develop an infrastructure for the persistent preservation of large amounts of data. Their approach is based on structuring data in order to make it both self-describing and independent of particular software programs as well as on the creation of data grids for the management of and access to digital resources.⁴⁰

⁴⁰ Information about SDSC work can be found at <http://www.sdsc.edu/NARA/Publications.html> and <http://www.sdsc.edu/PAT/>; Internet; accessed 30 August 2004.

Section 3 Components of A New State Repository Program: Fundamental Questions, Models for Consideration, and Recommendations

The current California depository program must address several specific issues with regard to creating a new digital state documents access and preservation program, operating within a larger information infrastructure for California state government. These issues are addressed in the sections below. Each section will raise fundamental questions that the program must answer in order to move forward. These fundamental questions are highlighted at the beginning of each section. Model practices for consideration in other nations, states and organizations have been developed to address these questions are discussed in order to provide potential models and starting points for California. Recommendations for answering these questions are highlighted.

Unfortunately, there is no “one size fits all” solution for creating a working, sustainable state documents access and preservation program. In fact, there are few models in existence at all, although there are numerous projects and actions that should be examined to inform those involved in program planning. The California government environment, however, has several unique characteristics that will impact the program development. These characteristics include the California economy, the ongoing budgetary problems within the state, the complexity and independent nature of its government agencies, and the scope and volume of digital information being produced in California. Because of these factors, California compares more to a national program than to many of the current states within the United States. Thus, many useful program elements can be found among national models such as those of Australia and the U.S. Government Printing Office.

Section 3.1 Legal Issues: Definitions, Authority and Compliance

Little of the language of the current Library Distribution Act (LDA) or the corresponding State Administrative Manual (SAM) guidelines for agencies addresses the digital publishing environment in California. The California Public Records Act defines “public records” as “any

writing containing information relating to the conduct of the public's business prepared, owned, used, or retained by any state or local agency regardless of physical form or characteristics,” and explicit states that “public records stored in a computer shall be disclosed as required by this chapter.” (California Government Code 6250-6270) There is little explicit recognition in California law and regulations, however, that public access to government electronic information should be assured on a permanent basis. California is not alone in the lack of legal language and updated guidelines to address these issues. A recent survey of the 50 states by the American Association of Law Libraries found that “very few states have updated their statutes to explicitly

FUNDAMENTAL QUESTIONS

- Is legal change necessary?
- What legal changes are appropriate?
- What about compliance and the law?

Managing and Sustaining a State Government Publications Program in California

incorporate electronic government information into their public access, freedom of information or depository laws.”⁴¹

Ultimately, legislative change will be necessary for California to create an effective program model for digital state publications. The State Library, however, has the authority to amend the State Administrative Manual to strengthen the depository program and to act to redesign the program to fit the digital publishing environment. Legislative change does not have to occur before other significant actions can take place.

The staff of the current California depository program frequently laments the lack of state agency compliance to the LDA. There are few examples of other state repository programs that have brought about legislative change to impose sanctions on state agencies for lack of compliance. In one instance, Arizona Revised Statutes section 41-1346 states that the head of any state or local agency that fails to comply with the provisions of the law regarding management of public records is guilty of a class 2 misdemeanor.⁴² Equally rare are programs that have engineered legal incentives to reward agencies for sending their publications to the appropriate organizations.

When incentive programs exist, in states such as Washington, Illinois, and Georgia, they are most often programmatic and service-oriented in nature, offering agencies training opportunities, meeting with appropriate agency staff to establish ongoing relationships, or offering agencies technical tools, such as metadata generators. This incentive approach was advocated by several of the March 2004 conference participants as having the most likelihood of success. The models below illustrate ideas for legislative mandates to address depository programs.

Models for Consideration

Colorado. Colorado has been recognized by the American Association of Law Libraries as a leader in enacting laws to clarify actions relating to the deposit of electronic state publications. Chapter 373, 22. 2003 (amending Colo. Rev. Stat. 4-90-204) defines publications as

“...including those made available through a public telecommunications network, an electronic copy or notification of the publication of such electronic copy shall be deposited with the center in a form specified by the center.” (Chapter 373, 22. 2003)

The law goes on to state that agencies must provide documents to the State Publications Depository and Distribution Center “within ten working days of publication.”⁴³ The law addresses permanent public access to state publications, directing the State Library to coordinate “with state agencies, depository libraries, or other entities [to provide] permanent public access

⁴¹ Richard J. Matthews, et al. *State-by-State Report on Permanent Public Access to Electronic Government Information* (Chicago, IL: American Association of Law Libraries, 2003), p. 3; available at http://www.il.georgetown.edu/aallwash/State_report.pdf; Internet; accessed on 30 August 2004.

⁴² *Arizona Revised Statutes*, section 41-1346. <http://www.dlapr.lib.az.us/records/l-gov.cfm>.

⁴³ Colorado. *Colorado Library Law. Part 2, section, 24-90-204, Deposits of state publications* (Denver: Colorado Dept. of Education); available at <http://www.cde.state.co.us/cdelib/LibraryLaw/Part2.htm>; Internet; accessed on 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

to state publications, regardless of format.” (Chapter 373, 23. 2003, amending Colo. Rev. Stat. 24-90-205)⁴⁴

Illinois. In Illinois, legislative changes have been made to the State Library Act to provide an incentive to agencies to deposit electronic publications. The incentive allows agencies to reduce the number of paper copies deposited if the document is also submitted in a specified electronic format.

“Sec. 21. Publications and lists; deposits by State agencies.

(a) All State agencies shall provide and deposit with the Illinois State Library sufficient copies of all publications issued by such State agencies for its collection and for exchange purposes. The State Librarian shall by rule or regulation specify the number of copies required and the publications that must be deposited. **The number of paper copies of a publication that must be deposited may be reduced if the documents are also submitted in an electric format specified by the Illinois State Library. The State Librarian shall set by rule the standard to be used for electronic data exchange among State agencies and the State Library.**

(b) For the purposes of this Section:

"State agencies" means every State office, official, department, division, section, unit, service, bureau, board, commission, committee, and subdivision thereof of all branches of the State government and which agencies expend appropriations of State funds.

"Publications" means any document, report, directory, bibliography, rule, regulation, newsletter, pamphlet, brochure, periodical or other printed material paid for in whole or in part by funds appropriated by the General Assembly or issued at the request of a State agency, excepting however, correspondence, inter-office memoranda, and confidential publications.

"Published material" means publications in print and electronic formats duplicated by any means of duplication, including material downloaded from a publicly accessible electronic network.

(c) The State Librarian shall from time to time provide a listing, electronically, in printed form, or in both formats, of the publications received by him or her under this Act.”⁴⁵ (emphasis added by the authors)

Georgia. In 1993, the Digital Library of Georgia was created through legislative actions, and the University of Georgia libraries assumed responsibility for Georgia government publications. In 2000, amendments to the Code of Georgia established the requirement for submission of electronic documents. Since that time, Susan Tuggle, Coordinator of Georgia Government Publications, reports that compliance has risen to over 50% of the agencies, in part due to an active outreach and education component. These efforts are supplemented by assistance provided to agencies for electronic submission.⁴⁶

The Georgia depository requirements are stated as:

“The University of Georgia Libraries was designated the official depository for Georgia government publications in 1993. As mandated by the Official Code of Georgia Annotated 20-5-2, state agencies and departments within the Georgia state government are required to submit five copies of all their

⁴⁴ Ibid.

⁴⁵ *Illinois Compiled Statutes, State Library Act* (15 ILCS 320/ from Ch. 128, par. 101).

⁴⁶ Susan Tuggle, Coordinator of Georgia Government Publications, University of Georgia Libraries, Interview, 11 March 2004.

Managing and Sustaining a State Government Publications Program in California

publications, including print, video recordings, CD-ROMs, and audio CD-ROMs to the following address: Georgia Documents Depository Collection, Government Documents Department, University of Georgia Libraries, Athens, GA 30602-1645.” Official Code of Georgia Annotated, 2000 Supplement, 20-5-2”⁴⁷

Washington. In 1998 the Washington State Legislature charged the State Library, State Archives and the Department of Information Services to create a system to increase public access to government information. The Washington Government Information Locator Service (WAGILS) and program was the result. In 2000 the WAGILS initiative became a permanently funded program of the State Library. Among the characteristics mandated for the program were:

- Ease of operation by citizens;
- Access through multiple technologies, such as direct dial and toll-free numbers, kiosks, and the Internet;
- Compatibility with private online services; and
- Capability of expanding the electronic public records included in the system.
- RCW 43.105.290

Among the charges to the State Library were:

- Promote and facilitate electronic access to public information and services;
- Establish content-related standards for common formats and agency indexes for state agency produced information;
- Providing advisory services to state agencies regarding their information needs.
- RCW 27.04.045⁴⁸

By 2001, the Washington Legislature recognized the need to address preservation of and permanent public access to government information and approved a plan to establish the first state government sponsored digital government archive repository with RCW 40.14.020.⁴⁹ The repository will open for business in fall 2004, with the Washington State Archives as the lead agency.⁵⁰

RECOMMENDATIONS FOR ACTION

Draft model legislation, drawing upon several concepts from the models stated above, including:

- The explicit acknowledgement that state government digital published information, including Web pages, is within the collecting scope of the new program.
- A mandate for the new program to provide permanent public access to, and preservation of, digital state government published information.

⁴⁷ *Official Code of Georgia Annotated. 2000 Supplement, 20-5-2.* Georgia depository requirements available at <http://www.libs.uga.edu/govdocs/collections/georgia.html> - information; Internet; accessed 30 August 2004.

⁴⁸ About the WAGILS program available at <http://find-it.wa.gov/gilsabot.htm>; Internet; accessed 30 August 2004.

⁴⁹ *Revised Code of Washington*, RCW 40.14.020.

⁵⁰ Cathy Turk and Adam Jensen, *Washington State Digital Archives Project Investment Plan* (Olympia, WA: Office of the Secretary of State, August 2003); available at http://www.secstate.wa.gov/archives/pdf/digital_archives/Investment_Plan.pdf; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

- An incentive to the state agencies, based on the Illinois model, so the agency distribution requirement is lessened if their publications are in digital form.
- A specific statement giving the new program authority to set the standards for data exchange between agencies and the program.
- Language specifically allowing the new program to establish content related standards for common formats, which would mandate creation of a metadata standard and/or a mark-up standard for publications.

Preferred Model and Outcome

This legislative change will ideally take place as part of planning for the legal components of the larger information infrastructure and in conjunction with the recommendation of the California Performance Review (CPR) report regarding state digital records. The report states:

The Governor should issue an Executive Order requiring all state agencies to alert the State Library of publication of digital documents, Web sites or other products that may be candidates for permanent public access through the State Library.

As part of the Executive Order, the State Library should be required to immediately implement an electronic submission process to help state agencies alert the library to vital digital publications, bulletins, documents and other key public records, until the state implements a permanent solution.⁵¹

The State Archives, CalRIM, the Office of State Publishing, the depository libraries, and possibly the state data centers should also review legislation in an effort to create a legislative mandate for an information infrastructure. The State Library has an explicit opportunity in light of the CPR recommendation to begin creating the new program to encompass digital publications.

Policy changes and guidelines will have the best chance for success if they are presented according to the general mandates that are provided by the Office of the Governor, the California Performance Review, and the priorities of the legislature. An opportunity exists to leverage the momentum and funding priorities of these bodies to help establish the program's vision and goals.

Section 3.2 Copyright Considerations

In California the issues of copyright compliance, intellectual property rights management, and the process for requesting permission to use information from state Web sites and publications

⁵¹ California Performance Review, "State Digital Records Vanishing," in *A Government of the People for A Change: Report of the California Performance Review* (Sacramento, CA: California Performance Review, August 2004), vol. 4, section GG 45.; available at <http://www.report.cpr.ca.gov/cprprt/issrec/gg/eff/gg45.htm>; Internet; accessed 28 August 2004. This section states: "Public access to California state government information erodes as thousands of digital documents vanish each year. As a result, the state is losing important historical documents. The State Library must identify and implement processes to collect and protect state government's digital records."

Managing and Sustaining a State Government Publications Program in California

are, as in nearly every other state, extremely opaque. There are no certain or uniform policy guidelines to assist librarians or archivists as they work through these issues. While conducting research for this report, the investigators could find no conclusive or consistent indicators that the state of California claims copyright for all government information. Further, the California Public Records Act does not explicitly state that public records are in the public domain, though it could be inferred to do so. Generally, when a government remains neutral, or when there is the absence of a claim of ownership, the provisions of the Federal Copyright Act apply to a publication. Thus, copyright ownership would appear to be in place for agencies per federal law absent any statement for public domain in California law.

There is no one California law or regulation that places published California government information in the public domain. However the California Legislature historically leans far more towards providing and protecting free public access to government information, rather than restricting it. The investigators assert that the California Attorney General and other legal experts will need to provide legally authoritative guidance on these matters.

In November 2000, the California State Auditor issued a comprehensive report for the legislature on the issue of the state's intellectual property assets, titled *State-Owned Intellectual Property: Opportunities Exist for the State to Improve Administration of Its Copyrights, Trademarks, Patents, and Trade Secrets*. While this report was written from a perspective of intellectual property use for economic gain, it gives a good picture of the current state of affairs with regard to copyright and intellectual property rights within California state government. It also raises the issue of the conflict between state ownership of intellectual property or agency ownership claims and the principle of open government as espoused by the California Public Records Act.⁵²

State Ownership As Addressed by California Agencies

Various California government agencies show statements of ownership and rights on their Web sites per the requirements of the federal Digital Millennium Copyright Act. There is inconsistent practice, however, regarding copyright and ownership claims. For example, this ownership statement appears on the "My California" state portal:

In general, information presented on this Web site, unless otherwise indicated, is considered in the public domain. It may be distributed or copied as permitted by law. However, the State does make use of copyrighted data (e.g., photographs) which may require additional permissions prior to your use. In order to use any information on this Web site not owned or created by the State, you must seek permission directly from the owning (or holding) sources. The State shall have the unlimited right to use for any purpose, free of any charge, all information submitted via this site except those submissions made under separate legal contract. The State

⁵² California Bureau of State Audits, *State-Owned Intellectual Property: Opportunities Exist for the State to Improve Administration of Its Copyrights, Trademarks, Patents, and Trade Secrets* (Sacramento, CA: California State Auditor, Bureau of State Audits, 2000); available at <http://www.bsa.ca.gov/bsa/pdfs/2000-110.pdf> Internet; accessed on 30 August 2004. A follow-up report is available at <http://www.bsa.ca.gov/bsa-htmls/pdfs/sr2002/2000-110.pdf> accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

shall be free to use, for any purpose, any ideas, concepts, or techniques contained in information provided through this site.⁵³

A similar statement with a one-sentence difference appears on the site of the State Public Works Board:

In general, information presented on this Web site, unless otherwise indicated, is considered in the public domain. It may be distributed or copied as permitted by law. However, the State does make use of copyrighted data (e.g., photographs) which may require additional permissions prior to your use. **Furthermore, the unique branding of the site and various official seals and marks may not be used without permission of the State.** In order to use any information on this Web site not owned or created by the State you must seek permission directly from the owning (or holding) sources. The State shall have the unlimited right to use for any purpose, free of any charge, all information submitted via this site except those submissions made under separate legal contract. The State shall be free to use, for any purpose, any ideas, concepts, or techniques contained in information provided through this site.⁵⁴ (emphasis added by authors)

Inconsistent examples of different practices, language and usage regarding permissions, ownership statements and copyright information appear on nearly every California agency Web site. This leads to confusion for Web site users and for information managers responsible for collecting and maintaining state documents.

Models for Consideration

South Dakota. In South Dakota, the state clearly asserts copyright ownership.⁵⁵ Some classes of material, such as public records, are defined by the law as being in the public domain. The definition of public records in South Dakota includes “(7) Intellectual property that is in the public domain.”⁵⁶

Library of Congress Mapping the Internet Electronic Resources Virtual Archive (MINERVA) Project. At the Library of Congress, the MINERVA Web archiving project copies Web sites for archival purposes. Project partners include the Internet Archive, Alexa Internet, and WebArchivist.org. Project staff catalog individual Web sites at the collection level using Machine-Readable Cataloging (MARC) and create Metadata Object Description Schema (MODS) item level records. The project asks permission to collect from each site they identify for inclusion in the project. Generally they get excellent response, with most sites indicating they are pleased to be asked.

⁵³ California, *Welcome to California: Conditions of Use: Ownership* [state portal] (Sacramento: State of California), available at http://www.ca.gov/state/portal/myca_homepage.jsp [scroll to bottom of page and click on “Conditions of Use”], accessed 30 August 2004.

⁵⁴ California State Public Works Board, *State Public Works Board, Conditions of Use* [Web site], available at <http://www.spwb.ca.gov/Conditions.htm>; Internet; accessed 13 July 2004.

⁵⁵ *South Dakota Statutes*, 2-16-8. “Copyrights of material -- Contract for use of state's copyright. The South Dakota Code Commission shall provide the material authorized for publication by § 2-16-6 will be copyrighted by the State of South Dakota, in the name of the State of South Dakota. The commission may contract with printers, publishers, and computer retrieval companies for use of the state's copyright.”

⁵⁶ *South Dakota Statutes*, 13-49-31.

Managing and Sustaining a State Government Publications Program in California

This process includes requesting permission from a site owner prior to copying the Web site for the archive. The permission request includes a notice of crawl, and request for permission to display the Web site offsite. It also includes an assertion by the Library of Congress of fair use under provisions of the U.S. Copyright code for event-based collection with citation of regulations, including section 407 of the U.S Copyright Code requiring government agencies to comply with mandatory deposit regulation.⁵⁷

Nevada. The state of Nevada provides more detailed information about their claims of ownership in the 2002 document, *A Web Style Guide*. Under the policies outlined in the document, although the content of executive branch Web sites is available to the public, certain information may be copyrighted, servicemarked or otherwise protected as the state agency's intellectual property. The policy provides for the following:

- All agency content is protected by federal copyright laws.
- Use of any intellectual property that was not created by Nevada state employees, or by contract for Nevada state use, must be in accordance with this policy pursuant to federal and state law.
- Appropriate credit or legal release from the content owner or copyright holder must be clearly visible.
- Examples of copyrighted content include news articles, graphics, photographs, copyrighted Web page design, animation and sound, etc.⁵⁸

Links to further examples of guidelines relating to intellectual property regulations used by some California agencies and for the states of Maine, South Dakota, Michigan, and Nevada appear in the copyright listing in Appendix 2 of this report.

RECOMMENDATIONS FOR ACTION

The agencies involved in planning the new program should obtain a legal opinion from the State Attorney General, or other appropriate counsel, as to the copyright status of government publications. This opinion should address several issues:

- Are California government publications under copyright protection? Who is the owner of the publications?
- Can the new program collect, preserve and provide access to electronic government publications if they are under copyright?

⁵⁷ Allene Hayes, "Web Archiving: New Developments." Presentation at the American Library Association Midwinter Conference, San Diego, California, January 11, 2004. The Library of Congress MINERVA project cites U.S. Code Section 407 as the applicable copyright law being used to allow archival access - see <http://www.law.cornell.edu/copyright/copyright.act.chapt4.html> - 17usc407; Internet; accessed 30 August 2004 for the text of the law. "Event-based collection" is defined as harvesting a Web site based on a schedule driven by an event. A certain level of change in a Web site (seen as an event) would prompt a new harvest or collection of the site. (Definition supplied by the investigators.)

⁵⁸ Nevada IT Operations Committee, e-Government Steering Committee, *Web Style Guide* (Carson City, NV: Nevada IT Operations Committee, 2002); available at [http://www.nitoc.nevada.gov/NITOCdocs/3.02\(A\)_Policy_WebStyleGuide.pdf](http://www.nitoc.nevada.gov/NITOCdocs/3.02(A)_Policy_WebStyleGuide.pdf); Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

- What actions does the new program need to take to ensure copyright compliance in the digital environment? Are there access issues, copying issues, etc. that need to be addressed?

The stakeholders will need to advocate and pursue a copyright environment in California that ensures that the new program can legally collect, provide access to, provide fair-use copies of, and take preservation actions upon state information, records, and publications regardless of format.

To the fullest extent possible, the program should request permission from state agencies to copy selected Web sites and documents. Such action will model appropriate rights and permissions action for agencies.

Preferred Model and Outcome

Further legislative action may be appropriate, depending on the opinion rendered. Copyright has generally not applied to public records; but copyright is being heavily contested and reconsidered currently and the general rule is “don’t assume anything.”

Section 3.3 Data Capture

Data capture is a broad term that reflects the many methods the new program can use to select and electronically acquire digital state publications. Data capture can serve multiple functions including access, retention, and preservation of documents and information. Because the California government environment is large and complex, several methods will be necessary in order to ensure as much content is collected and retained as possible. Since the ultimate purpose of these activities, such as cataloging and indexing, are done to support public access, they should be conducted at all times with the end-user in mind.

FUNDAMENTAL QUESTIONS

- How do we acquire government publications?
- What publications should be collected?
- How can we work with agencies to facilitate acquisition and management?
- How will we describe and manage digital publications?

Traditional library and archival theory and practice are based on an underlying principle that the creators of government records and information are not the best long-term custodians of their resources. As Graham Toft states, government management is near-term and action oriented, and does not tend to focus on the long term.⁵⁹ Libraries and archives are specially equipped to facilitate and sustain access and preservation. Additionally, they can ensure that the information in the library or archives is authentic, reliable (it hasn’t been tampered with) and is not destroyed as a result of political motives.

⁵⁹ Graham Toft, “Synoptic approaches of strategic management,” in *Handbook of Strategic Management* (New York: Marcel Dekker, c2000), p. 5.

Managing and Sustaining a State Government Publications Program in California

In the digital environment, this traditional model continues to be the best option for materials of enduring value. Yet environmental realities suggest that, at least in some cases, agencies may manage digital resources for a relatively long period of time, and only eventually transfer them to the library or archival program for permanent preservation and access. We will discuss data storage and management issues later in Section 3.4.

The issues with data capture (the identification and collection of digital publications) are vital to the development of a long-term program for access and preservation.

With this in mind, this section approaches data capture based on these core tenets:

- Data capture serves several purposes, including access, retention, and preservation.
- The new depository program will obtain physical custody of select digital publications based on established selection criteria. Content selection is addressed in part 3.3b of this report.
- Acquiring, managing and providing access to content should be the focus of the program.

Data capture methodologies will need to be diverse in order to meet the needs of the state agencies. Some agencies may actively participate in the new program, while others may not participate at all.

Core content that is particularly valuable must be prioritized. Legislative bills and the governor's budget, for example, are highly important documents vital to both history and the ongoing functioning of government. Brochures or minor reports are not as vital. The new program, particularly in the beginning stages, must make these types of, often difficult, selection decisions.

It will be important to create an inventory of the ca.gov domain at some time early in the development of the new program. This will allow the program to quantify the size and scope of the digital and analog publishing environment in California government. It is necessary to set the business requirements for any technological approaches that are developed. Understanding the scope, size and depth of the problem will help the steering committee properly match the available resources with the tasks to be done and will give the program's leadership an understanding of the pace of activities and phases of planning. Finally, when the size and scope of the problem is better understood, a more compelling case can be made for obtaining funding from the legislature and granting organizations. Models for determining and understanding the size of the ca.gov domain are currently in use by the California Digital Library,⁶⁰ the Find-It Illinois research model,⁶¹ and the inventory model of the North Carolina Access to State Government Information initiative.⁶²

⁶⁰ Patricia Cruse, et al. op cit.

⁶¹ Larry Jackson, *Preserving Electronic Publications: Final Report* (Springfield : Illinois State Library, 2002) available at http://www.cyberdriveillinois.com/library/isl/lat/pep/pep_final.pdf; Internet; accessed 11 May 2004.

⁶² Kristin Martin and Jan Reagan, *North Carolina State Government Information: Realities and Possibilities* (Raleigh, NC: State Library of North Carolina, November 2003); available at <http://statelibrary.dcr.state.nc.us/digidocs/Workgroup/WhitePaper.pdf>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

Section 3.3a The Role of Content Producers (State Agencies) in Data Capture

State agencies are the creators of government publications. Under the Library Distribution Act, their role is to ensure that their publications are distributed to the State Library and depository libraries. The entire system is based on agencies fulfilling this role; their failure to do so has already been established. When they fail to do so, the library distribution system cannot function. With this in mind, and in consideration of the current digital environment, what is an appropriate role for the state agencies in the new program? The lack of communication and interaction between the agencies and the depository libraries has already been established.

The roles and responsibilities of the state agencies and data centers within the new program must be clearly defined and recognized. The leadership role of the State Library has been recognized in the recommendation of the California Performance Review.⁶³ The process of identifying and describing roles and responsibilities for state agencies, the data centers, the State Archives, State Library, and the other state agencies involved in access and preservation of state government information can be assisted by the development of pilot projects to test some new methods for electronic deposit of agency publications. Among the most successful and mature programs, including Illinois, Australia, and the GPO, an active climate of testing and demonstration of models is underway. The investigators assert that an environment of collaboration, with partnerships based on clear benefit to both the depository libraries and the agency creators will provide the best incentive for development of a new depository library program. With this in mind, what is an appropriate role for the state agencies in the new program?

The collaboration of the agency content producers will be essential to the success of the new depository program. Further, the development of a true information infrastructure for California government will require a more active role for both the information managers and the information producers. As active partners, the agencies may take on some tasks, such as managing their own agency content as part of a network of repositories, describing information content according to established standards and becoming leaders in developing policies, standards and processes.

Some state agencies and programs are already committed to managing their own resources. These include the Legislative Analyst's Office, the Commission on Peace Officer Standards and Training, and the California Employee Development Department. At the March 2004 conference, representatives of several agencies, including the California Environment Resources Evaluation System (CERES), the Health and Human Services Data Center, the Legislative Data Center, the California Department of Water Resources, and the Corporation for Education Network Initiatives in California (CENIC), expressed interest in participating in a new program to preserve and provide persistent access to California digital government information. They stressed the need for specific guidelines and standards to assist them in working with the program leadership and other stakeholders.⁶⁴ Conference participants also mentioned the need for public relations, regular communication with agencies, training and education.

⁶³ California Performance Review, op cit.

⁶⁴ See Appendix 1 of this report, "Final Report, California State Library Government Documents Conference," p. 17-18.

Managing and Sustaining a State Government Publications Program in California

Under the new program, meaningful agency compliance incentives can only be identified through an active dialogue with agencies themselves. The dialogue can most productively occur if agency representatives serve in leadership roles on the steering committee that plans the new program. During interviews with some agency staff, the investigators found evidence that agencies are interested in finding ways to assure that the information and publications they work hard to produce can be accessed by California inhabitants over the long-term. The ideas for incentives offered during the conference included:

- Make it easy for the agencies; give them instructions.
- Notification/push databases by agencies to depository libraries.
- Metadata tools provided for the agency by program leadership.
- Give awards—superstar, most improved agency site.
- Educate agencies on how they can save money by having a repository program/libraries take care of storage.
- Educate agencies on how a repository program can benefit them in terms of payoff for agency security (cataloging and backup).⁶⁵

Some representatives from agency programs have already volunteered to work toward developing a new repository program because they see the benefit to their programs in doing so. During the March 2004 conference, individuals from ten state programs volunteered to serve either on a steering committee or to serve on a working group for a particular area of investigation. Representatives from four state programs, including the State Archives, the Department of Water Resources, the Department of Parks and Creation, and the Department of Fish and Game, also volunteered to participate in one or more test bed projects to investigate potential technical solutions. A complete list of individuals from state agencies and programs who have volunteered to participate in developing a new repository program appears in Appendix 1.

Models for Consideration

There are no examples of significantly new models in this area. State programs that have successfully advocated for legislative change to preserve and provide enduring access to digital government information continue to base their programs on voluntary compliance by state agencies. Generally these programs are supported by education and training activities or based on active outreach activities as incentives for that compliance.

Georgia, Washington, Illinois, and Texas. The depository programs in these states have worked with agencies using active communication, education, and outreach to raise awareness of the depository mandate and generate interest in the depository program. These activities have included marketing and training components.

Most of these depository programs include a set of guidelines for participation, including metadata standards for agencies to embed descriptors in their publications so they can be captured and indexed. Clear instructions for participation are published online and reinforced

⁶⁵ Ibid, p. 15-16.

Managing and Sustaining a State Government Publications Program in California

with active marketing. Defining a partnership model with agencies and delineating the roles and responsibilities for agency participation has had a positive affect on communication for these states. After state documents librarians in Georgia published specific electronic repository guidelines and outlined responsibilities for participation in their program, a 50% increase in compliance in depositing both print and electronic documents was noted.⁶⁶

Many advocate this approach, yet given the size and complexity of California government it is highly unrealistic to expect that this outreach-oriented approach will be completely effective as a sole solution. There must also be an active data capture component.

Some states, including Georgia, Washington and Texas, have had success in convincing state agencies to provide alerts about digital publications as they are published. Convincing agencies to proactively alert the depository program about available analog and digital materials has required marketing, promotion and training. In these states, participation requirements for the electronic depository program are published and promoted to agencies. The same is true of the program for the National Library of Australia.

Arizona. Arizona is beginning to address this issue from the opposite direction. They assume that the agencies are not going to help the State Library and thus the Library must acquire digital state publications without agency assistance. In this developing program, the state agencies' role disappears and the Library takes on all responsibility for capture.

All of these models use or have used harvesting technology to capture digital publications from agency Web sites. Currently there are two types of harvesting being conducted:

Selected Harvesting

This human-based selection mimics traditional library workflow and requires staff devoted to selection and harvest. In large environments, such as California, the number of Web based publications is so large that a human-based selection and harvest mechanism, if it is the only tool used, will not be adequate to meet the large volume of annual publication and distribution.

Bulk Harvesting

The second kind of tool also harvests from Web sites, but it is a larger harvest that captures entire Web sites, or designated portions thereof. Rules can be written to define and limit the automatic harvesting. This more automated method of selection and harvesting has the potential for becoming a more satisfactory method because it requires less human intervention – letting the computers do the work as much as possible. However, to make the harvest results truly useful, these bulk harvesting tools still require staffing for selection and management and to refine and direct the harvested information into search and retrieval systems amenable to citizen access.

Automated bulk harvesting of government data is a commonly utilized method of data capture because it is relatively easy to accomplish and there are many tools available to support it. As such, it may be an appropriate starting place for the new program and will most likely be a

⁶⁶ Susan Tuggle, op cit.

Managing and Sustaining a State Government Publications Program in California

method used for some digital content for an extended period of time. Once acquired, the data can be sorted, indexed, cataloged, described using metadata, and classified for current access purposes. The same harvested data can be sorted and selected for further processing for such purposes as permanent retention and permanent public access. These second level processing activities are more difficult, take more human intervention and will require additional highly specialized tool sets to achieve intended results. In the new repository program partnerships with several key stakeholder agencies will be required to accomplish all the steps and tasks involved to produce maximum value from data capture activities.

Manual harvesting allows the library staff to control what is collected, but is time consuming and labor intensive. Bulk harvesting appears to provide less-than-satisfactory control over what is collected, but is faster and less labor intensive. Neither solution adequately addresses the issue. Further, harvesting does not effectively gather content from the “deep Web” – the underlying databases and systems that are the source of much Web content.

RECOMMENDATIONS FOR ACTION

The role of state agencies should be clearly defined and developed through agency participation in the planning process for the new program. The new program should incorporate multiple models for data capture and management.

State agencies should be categorized based on their ability and willingness to comply with the distribution mandate. Based on this categorization, the new program should provide agencies with the tools appropriate for their category. Categories could include:

Repository. An agency has an active program for the ongoing storage, management of, and access to their state publications and has an agreement with the new repository program that their system meets defined requirements. If the agency is unwilling to hold content indefinitely, the agreement should also state when the agency would transfer materials to the repository.

Active Notification. An agency actively notifies the repository about new publications via an alert mechanism that could be as simple as an email message. Upon notification, the publication should be manually or automatically harvested.

Active Distribution. An agency actively sends copies of publications to the repository via print, CD, ftp, or other data transfer protocol.

Automatic Harvest. An agency is not actively participating in the management, notification or distribution of publications. The repository will have to rely solely on harvest techniques to capture its publications. In order to successfully harvest output from agencies’ Web sites and access portals, the host agency will need to be aware of the process.

The new repository program should actively conduct education and outreach initiatives for state agencies and enlist active participation of agencies in the effort to identify, capture, maintain and preserve government information.

Managing and Sustaining a State Government Publications Program in California

Preferred Outcomes

Outreach and education will not be effective for every agency in California. Agencies should be targeted based on their willingness to participate and the value of the publications they create. Outreach should include a proactive campaign to raise awareness of the goals and objectives of the new program. Additionally, state agencies should be active participants and leaders in developing the new program.

Section 3.3b Content Selection

Other issues aside, a fundamental question for a new repository program for preservation and persistent access to California state government documents is, “What should be collected?” Content selection should not be based on the capabilities of particular software, but should be based on the nature and content of the publications. Ultimately, content selection criteria for digital publications should be no different than for paper-based publications. Yet the digital environment necessitates a more proactive role for the new program. No longer can the program rely on agencies to send publications; instead priorities must be set for the active solicitation of content from the agencies. With limited resources, establishing these “priority publications” is an important decision-making process.

As we noted in the earlier section on legal issues, most state documents programs narrowly define their collecting scope to distinct, stand-alone digital publications, similar to paper-based publications. This practice calls for us to ask the question, “What about entire Web sites?” Government and government-related Web sites are volatile, particularly during times of administration change. California must address both the content of distinct publications and of Web sites in order to effectively document the activities of government. The argument can be made that a Web site, because of its public nature, is in and of itself a publication. However, an entire Web site is much more analogous to an archival collection or record group, which contains many types of information and records, than it is to a publication.

Models for Consideration

National Archives and Records Administration/ U.S. Government Printing Office. While the GPO collects and provides access mostly to distinct publications, the U.S. Federal Records Act requires agencies to identify and transfer Web site records to NARA for permanent retention.⁶⁷ In late 2000 NARA led an initiative to capture all Web sites at the end of the Clinton administration. NARA has also worked with the San Diego Supercomputer Center on a project investigating the preservation of presidential Web sites.⁶⁸

Library of Congress. The Library of Congress works with other organizations in a collaborative effort to meet these challenges. Working with the Internet Archive, WebArchivist.org and the

⁶⁷ U.S. Code, Title 44, Public Printing and Documents, Chapter 31, Records Management by Federal Agencies; available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1994_uscode_suppl_3&docid=44usc3101; Internet; accessed 30 August 2004.

⁶⁸ Amarnath Gupta, *Preserving Presidential Library Websites* (San Diego, CA: San Diego Supercomputer Center, 2001); available at <http://www.sdsc.edu/TR/TR-2001-03.pdf>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

Pew Internet & American Life Project, the partners have created a collection of digital materials from the Web known as the *September 11 Web Archive*.⁶⁹

Arizona. Arizona is beginning to conceptualize a framework for a new approach to content selection, based on an archival theory called macro-appraisal. In this approach, Web resources are evaluated based on provenance and informational content, called macro-appraisal. Inherent in this idea of macro-appraisal is the recognition that some information is more important than others, and some agencies are more important than others. For example, the Governor's Web site and publications are more important to the ongoing documentation of government than the (fictitious) "Barber and Hair Dressers Board." Also incorporated into the concept is the volatility of the Web site (some digital documents will need to be selected more often than others, due to their high volatility.) This approach acknowledges that 1) not everything can be captured with existing tools and resources; 2) some digital publications (and records) are more important to capture than others.

RECOMMENDATIONS FOR ACTION

The new program should create and implement technology-independent selection criteria based on macro-appraisal decisions. Technical tools should support the selection criteria.

Preferred Outcomes

The creation of selection criteria, based first on provenance (creating agency) and second on content. This could result in a ranking of agencies that indicates which agencies have content that the program: a) must have (priority publications); b) would like to have; c) would take if offered; or d) doesn't want. Based on this ranking, the new program would contact the "must have" agencies and determine the best way to work with them based on the agency role categories outlined earlier.

To ensure adequate collecting, a decision must be made regarding the selection, preservation and ongoing management of entire Web sites. It seems unlikely that the entire ca.gov domain could be captured and preserved in its entirety, every time a change occurs, *and* that users would also be able to easily identify needed distinct digital publications. This effort would be massive and, at the present time, is unrealistic both technically and managerially.

More likely, the new program will be able to capture, store, manage, and provide access to snapshots of a majority of the ca.gov domain. Thus the new program would essentially maintain a picture of ca.gov at several points in time – possibly quarterly or annually. This effort is feasible and would provide a "sampling" of the domain for research purposes. It would not, however, be all-inclusive. A second component of the program would provide immediate public access to distinct digital resources.

The new repository program would take responsibility for state government Web site selection and capture, and for ongoing preservation and access to those resources. This function would be

⁶⁹ Library of Congress, et al., *September 11 Web Archive* [Web site] (Washington, DC: Library of Congress MINERVA Project); available at <http://september11.archive.org>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

performed separately from the function that would provide immediate public access to digital state government publications. Both functions would provide a good long-term view of the ca.gov domain and would ensure that digital records and Web sites as well as digital publications are available persistently.

Distributed content selection is investigated and tested appropriately to ensure its viability. An effective distributed model should result in all of the “must have” resources being collected and managed and a significant percentage of the “would like to have” resources being collected and managed. Depository libraries or related programs would be considered as potential partners in this process.

Section 3.3c Metadata

Metadata is the data that describes the characteristics of a digital resource. Metadata may be descriptive, providing access to the content; technical, providing information about the technical nature of the content; administrative, providing the information necessary to effectively manage the resource; or structural, providing information about the internal structure of the resource and the relationships among the files that make up the resource.

California state government does not currently have a fully developed standard descriptive metadata set for use by government agencies. There are suggested metadata fields that are recommended as part of the style implementation guidelines published for the California state portal.⁷⁰ Additionally, the library community has guidelines for applying metadata as part of its LSTA funding for digitization projects for the State Library, available at <http://www.library.ca.gov/assets/acrobat/metadocfinalrev.PDF>

It is not clear that having a metadata standard for agencies to use as they create Web resources is effective, particularly without adequate training and outreach. States (including Washington, Illinois, Minnesota, and Ohio) that have implemented such standards to support agency generated metadata often find that compliance is low, even when tools are provided to help with the creation of the metadata. Metadata will need to be created for digital resources, however, whether done by the agencies or not, and defining standard metadata sets and schemas will be important to the development of the new program. It may also act as a starting point for developing metadata standards for the information infrastructure.

Models for Consideration

Standardized Descriptive Metadata. Numerous states have developed standard descriptive metadata sets for agencies. These include Washington, Illinois, Minnesota, and Ohio. The general trend is to use the Dublin Core metadata set, and provide guidelines and tools for creating metadata. The Dublin Core metadata set is a National Information Standards Organization (NISO) standard.⁷¹

⁷⁰ California. *Style Implementation Guidelines, HTML guide* (Sacramento: Teale Data Center); available at <http://www.webmasters.ca.gov/styleguide/META>; Internet; accessed 30 August 2004.

⁷¹ National Information Standards Organization, *The Dublin Core Metadata Element Set* (Bethesda, MD: National Information Standards Organization, 2001); available at <http://www.niso.org/standards/resources/Z39-85.pdf>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

Structural Metadata. The National Archives and Records Administration is using structural metadata, via their work with the San Diego Supercomputer Center. The model uses Extensible Markup Language (XML) as a method for the creation of self-describing documents. XML is being implemented widely in library and archival communities to create structured data and provides an excellent tool for sustained data management.

Technical and Administrative Metadata. Technical and administrative metadata are dependent on the context of the program. Technical metadata elements will depend on the type of file formats being managed. Currently the only standard technical metadata set under development is for digital still images. The Data Dictionary - Technical Metadata for Digital Still Images (NISO)⁷² remains in draft form at the time of this report.

RECOMMENDATIONS FOR ACTION

A descriptive metadata standard, based on the NISO standard Dublin Core metadata set, should be created and promulgated by the new repository program and ultimately for the information infrastructure.⁷³

Preferred Outcomes

The descriptive metadata set should be the mandatory core set for agencies acting as repositories as well as for stakeholder organizations that are managing data in order to support data exchange, interoperability, and access. The set should also be promoted among state agencies. Ideally, content creators will include minimum metadata elements (title, author, date of publication) within the publications they create, which can be augmented with further metadata as necessary. However, it is realistic to expect that not all content creators will include metadata, and in these cases metadata will need to be created by the repository.

Structural metadata, as implemented via XML mark up of digital resources, should be implemented as widely as possible. Every effort should be made to ensure that “must have” resources are self-describing documents, created via XML.

Agencies create their textual information resources using XML, thereby creating self-describing and non-proprietary documents that are easier to manage over the long term and which already contain descriptive metadata.

Technical metadata are a function of data management and digital preservation. They should be considered within that context when it is appropriate within the new repository program development.

⁷² National Information Standards Organization, *Data Dictionary – Technical Metadata for Digital Still Images* [draft standard] (Bethesda, MD: National Information Standards Organization, 2003); available at http://www.niso.org/committees/committee_au.html Internet; accessed 30 August 2004.

Section 3.4 Data Management and Access

Permanent public access implies ongoing data management in order to facilitate access via trustworthy repositories. It encompasses a variety of components, including system design, security, access, administrative metadata, network connections, etc. Data management is the nuts and bolts of the new program and will require significant input from technical experts and computer scientists as well as librarians and archivists.

FUNDAMENTAL QUESTIONS

- When should end users be considered?
- How should our repository be structured?
- How can a repository ensure that publications are authentic?

David Harris, Internet Services Coordinator for the California Environmental Resources Evaluation System (CERES), wrote in his evaluation of the March 2004 conference:

One of the most achievable ideas from the event was to form a workgroup to pursue implementation. The level of expertise and commitment was exceptionally high. (I'm most enthused about) the opportunity to put more of the non-GIS ideas developed at CERES into practice over a much broader subject matter. While it's far from trivial, I believe it is possible to develop distributed tools to support improvement of access to local information assets within departments which would use standard indexing, data structure and communications methods providing a local benefit which would provide an incentive to organize information assets so that they are accessible across all departments.⁷⁴

It is beyond the scope of this report to make specific recommendations for technology and tools that are appropriate for the new program. There are, however, several high-level considerations that are vital to sustainable data management and access that will be discussed here.

Section 3.4a Access

Prior to discussing the technical tools and methods that must be implemented to support permanent public access to and preservation of digital state publications, the role of the end-user of (and ultimate audience for) all these technological efforts will need to be examined. A fully developed and comprehensive program must be based on a user-centered approach where both short-term and persistent access are developed in integrated and complementary functions.

In this report we advocate that the new program recognize the different functions for preservation and access to entire Web sites and to distinct digital publications. The technology (systems, servers, storage arrays, etc.) that supports these two functions may or may not be separate. It is reasonable to imagine that the two functions could and would be interoperable and utilize the same technology.

⁷⁴ See "Evaluation Responses" in Appendix 1 of this report.

Managing and Sustaining a State Government Publications Program in California

Models for Consideration

Among a number of states that are actively involved in creating digital publications programs, the investigators found that there was interest in designing systems that would recognize the life cycle of materials. Both Arizona and Illinois have looked at using the initial harvest of government information from Web sites to feed both a current access portal and a digital preservation implementation. The North Carolina Access to State Government Information initiative recognizes the importance of users by including representatives from the public in its planning process.

RECOMMENDATIONS FOR ACTION

End users can provide useful input into discussions and planning for access to the digital repositories. Interested parties such as journalists, public advocacy groups and “good government” proponents are among the end-user groups that can provide comments about access models designed to meet the needs of the end-users.

The new repository program should investigate the merits of providing end-user access to digital state government publications through an established portal, such as the existing state portal, My California, versus establishing access through other systems, such as library catalogs.

Section 3.4b Repository Architecture

Most discussion of repository architecture distinguishes between centralized repository and distributed repositories. A centralized repository implies a single repository managed by a single organization; a distributed repository implies several repositories with several managing agencies and an oversight program to define standards and requirements. A repository is committed to ongoing management and access to resources, but not necessarily permanent preservation.

Repositories will act as main storage and access mechanisms; they may be (relatively) short-term repositories that eventually send holdings to another repository or long-term repositories (that assure permanent public access).

A key decision point is whether California should implement a centralized repository system or a distributed repository system. Either system is valid. The question is, which is more suitable for California? That question will be answered as the stakeholders come together, assess their resources, and design the new program.

Among the key activities for the new repository program will be to investigate and test new technologies and strategies for ensuring the long-term access to and preservation of digital collections in a cost effective manner. Technological models for acquisition, management, and preservation of electronic documents will need to be developed, tried, and tested. Technologies must support necessary workflow if the program is to successfully acquire and manage state government Web-based resources.

Managing and Sustaining a State Government Publications Program in California

The library and archival communities are in the early stages of developing new technologies that adequately reflect “real-life” workflows and require a limited amount of staff intervention. With so few existing programs that are implementing ongoing digital management and preservation, there is a true gap of available models and research in this area. In California, the San Diego Supercomputer Center is working on these issues and has several documents available regarding their Storage Resource Broker and data grid technologies, and the use of XML. Tools such as DSpace, which will be discussed later, may also be options for repository management.

Models for Consideration

Centralized Repository

- Physically held and managed by one organization.
- Holds all digital documents.
- Provides access to all digital documents.
- Requires a large commitment of technical expertise and tools from the managing organization (most likely coordinated by the State Library with actual management provided through partnership with the new consolidated state data center).

Distributed Repository

- Several repositories physically located in different places.
- All repositories meet defined requirements.
- Access provided via one-stop portal based on metadata from all repositories.
- Technical expertise and tools are distributed.

Cornell's Project Prism: Prism is investigating the use of risk management methodologies to create a “noncustodial, distributed model for archiving, in which resources are managed along a spectrum, from, at the highest level, a formal repository to, at the lowest level, the unmanaged Web.”⁷⁵ Prism is based on the idea that libraries and archives will increasingly be called to manage and provide access to content that is not in their physical control. Further, Prism concedes that some content providers will cooperate and some will not, and advocates a new model for handling this fact.⁷⁶

RECOMMENDATIONS FOR ACTION

Decide on an overall repository architecture for the new program.

⁷⁵ Anne Kenney et al. , “Preservation Risk Management for Web Resources Virtual Remote Control in Cornell's Project Prism,” *D-Lib Magazine* (January 2002); available at <http://www.dlib.org/dlib/january02/kenney/01kenney.html> - 19; Internet; accessed 30 August 2004.

⁷⁶ Cornell's Project Prism is available at <http://www.prism.cornell.edu/>; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

Preferred Outcome

The investigators believe that a distributed repository system is the best model for implementation. A distributed model promotes collaboration and uses the existing resources, interest, and expertise in the California environment.

California will have to be creative in developing this component of the new depository program. Expertise will be vital and should be available from the state data centers, the California Digital Library and the San Diego Supercomputer Center.

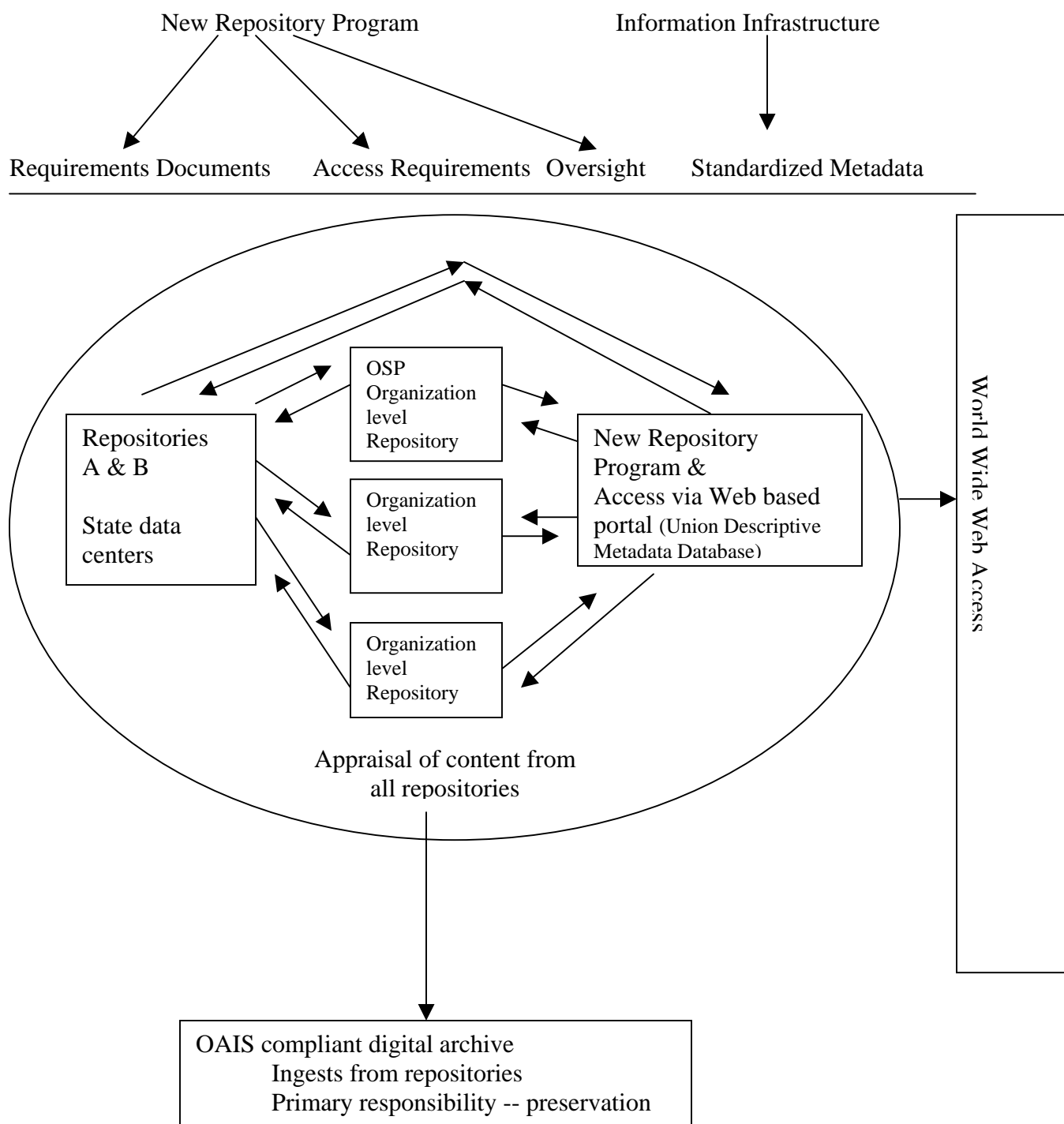
Requirements for interoperable repositories must be established and should be based on open system architecture and standards that comply with oversight authority under California law.

Should a distributed repository system not be possible, California should thoroughly investigate, evaluate and, if appropriate, implement the concepts and tools that should emerge from the Prism project.

The investigators can envision a high level structure for implementing data management that is included below (see Figure 2), but implementation must rest with the technical experts and reflect the political realities. This diagram is a functional model and could be implemented in a number of ways. Another such structure was outlined by James Jacobs of the University of California San Diego Libraries at the March 2004 conference. A rough sketch of this structure appears in Appendix 1 of this report.⁷⁷

⁷⁷ See Appendix 1 to this report, Attachment 2, "Follow-up Investigation."

Managing and Sustaining a State Government Publications Program in California



Repository = ongoing management and public access via defined requirements

Organizational Level Repositories = local repositories managed by state agencies and program partners, including depository libraries

Digital Archive = ongoing preservation and management activities

Figure 2

Managing and Sustaining a State Government Publications Program in California

Section 3.4c Authenticity

The term “authenticity” as used here addresses what a digital resource purports to be and how it was created. The authenticity of a digital object refers to the degree of confidence a user can have that the object is the same as that expected based on a prior reference, or that it is what it claims to be. An authentic digital record or publication is one that is genuine, that has not been counterfeited or tampered with, and is free from corruption.

Data management and repository implementation must be based upon trusted systems, where authenticity of documents and records is assured. Citizens deserve mediated and unfettered access via the World Wide Web. They desire free, user-friendly, durable, reliable, authentic, and ongoing access for California inhabitants. Ensuring that the digital information is authentic should be done with specific technologies and system rules that verify the integrity of the data and assure system security.

Models for Consideration

The investigators found few practical descriptions or examples of projects that have dealt with authentication of electronic documents. One notable exception is a report issued by the Amsterdam Municipal Records Office, describing its efforts to assess and authenticate a group of electronic records received on deposit. The records included a variety of electronic text formats contained on CD-ROMs, Web sites and databases. The author, Marcel van Dijk, concludes the article with this observation, "This project shows in essence one thing very clearly: when archiving of electronic records is not properly planned and projected right from the start, as with a proper administration of the contextual, administrative and technical metadata, then any reconstruction afterwards is bound to be a most laborious and time-consuming affair with an uncertain outcome. The chances of success decrease with the passing of time. Good procedures and agreements with record creators beforehand together with a strict 'admission policy' are indispensable to ultimate success."⁷⁸

Further guidance on this issue is offered in three sources:

- The January 22, 2004 report, *GPO Depository Library Council's Advice to the Public Printer*.⁷⁹ In this report, the council addresses version/authenticity control and digital information storage and legacy collections. Another useful discussion point in this source concerns the principal of assuring citizens continue to have the right of no-fee access to information by and about their government.
- The Council on Library and Information Resources May 2000 report, *Authenticity in a Digital Environment*.⁸⁰ This compilation of reports by different authors offers a

⁷⁸ Marcel Van Dijk, "It Always Hurts the First Time: Experiences with Transferred Electronic Records," *Cultivate Interactive*, Issue 9 (February 2003); available at <http://www.cultivate-int.org/issue9/amsterdammro>; Internet; accessed 30 August 2004.

⁷⁹ U.S. Depository Library Council, *Depository Library Council's Advice to the Public Printer* (Washington, DC: U.S. Government Printing Office, January 22, 2004); available at http://www.access.gpo.gov/su_docs/fdlp/council/012204_council_rpt.pdf; Internet; accessed 30 August 2004.

⁸⁰ *Authenticity in a Digital Environment* (Washington, DC: Council on Library and Information Resources, 2000); available at <http://www.clir.org/pubs/abstract/pub92abst.html>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

comprehensive discussion of the scope of issues involved in authentication of digital and electronic records.

- InterPARES. Among the organizations that are actively working on the issue of authentication is the international archival project, the International Research on Permanent Authentic Records in Electronic Systems (InterPARES). InterPARES aims at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form”.⁸¹

RECOMMENDATIONS FOR ACTION

A working group should be charged with developing tools and workflows that address the issue of authentication of materials as part of the repository requirements. State agency representatives should be part of the group in order to ensure they can meet and accept the standard.

Section 3.4d File Formats

File formats are relatively limited in the government domain. The HTML and PDF formats are the most widely used. That said, although many digital repository programs would like to be able to limit the file formats of the content they acquire, most accept that this kind of limitation isn't practical. However, it is possible for repositories to transform digital publications into standard file formats. It may be necessary to employ this practice in order to create and manage a cost-effective repository. TIFF, PDF (and soon PDF-A) as well as XML are generally considered as the “best bet” file formats to use for long term sustainability. HTML is a close second. There are no official or working standards yet for audio and video files.

One example of a model that accepts all formats is the Washington State Digital Archive facility. This repository, scheduled to open in fall 2004, will accept all file formats that agencies produce and submit. The archive plans to bring in the data, manipulate the data and re-format it into more standardized, XML-structured data for permanent retention. This model is experimental at present and is based on the previously mentioned work by the National Archives and Record Administration and the San Diego Supercomputer Center.

RECOMMENDATIONS FOR ACTION

Long-term data management will be best achieved if file formats meet standardized requirements as a condition of being accepted into a repository. The best file formats for long term sustainability are non-proprietary formats that support robust metadata, either embedded within the file or as mark-up, and provide the most potential for format migratability.

⁸¹ International Research on Permanent Authentic Records in Electronic Systems, (InterPARES), *InterPares 2 Project: The International Research on Permanent Authentic Records in Electronic Systems* [Web site]; available at <http://www.interpares.org/>; Internet; accessed 30 August 2004.

Section 3.5 Digital Preservation

The question of how to balance preservation and accessibility/usability must be addressed.

Implementing a digital preservation component will be a huge task for the new repository program. Core issues and approaches to digital preservation will be addressed here.

Determining how to preserve digital materials continues to be a challenge for the larger library and archival community. Although the development of standards and strategies is underway, all are in their infancy and require significant testing, improvement and proof of concept before they are truly reliable. Moreover, digital preservation has yet to truly be incorporated into most library and archival organizations in terms of staffing, funding, education, and infrastructure.

FUNDAMENTAL QUESTIONS

- What are the best models for implementing a digital archive?
- What is the OAIS Reference Model and why is it important?

The goal of digital preservation is to sustain the digital resource through time without losing significant informational and contextual properties. Preservation is the foremost concern, but access and usability must be ensured. Preservation may be achieved by creating images of digital resources using image formats such as TIFF or PDF. Such formats will allow for accessibility, but the usability of the data can be impacted. For example, a PDF of a report allows a user to see what the document looked like and even search within it; but will not allow for data manipulation. Ideally, long-term preservation will allow us to use digital data in new ways: to consolidate reports, to see long-term trends, to mine for metadata that is inherent within a document. The use of XML to create self-describing documents facilitates this ideal by using embedded metadata to describe the elements of a document, providing the ability to manipulate the elements via a database or other system.

Models for Consideration

LOCKSS. Lots Of Copies Keep Stuff Safe (LOCKSS) is a Stanford-based project originally designed to create caches, or copies, of digitally published academic journals on local servers and to automatically add new content. The LOCKSS staff has also worked on the viability of implementing LOCKSS for the Federal Depository Library Program. Their 2003 report states, “It is unclear whether the LOCKSS journal model – in which a single plug-in is associated with an individual journal publisher platform – could directly be applied to Federal agency content.”⁸²

The LOCKSS concept provides interesting new ways of considering the roles of the depository libraries, even if recent work indicates that the LOCKSS software itself may not be appropriate for government documents. Would the depository libraries consider acting as a repository for content that is immediately accessible to California inhabitants? Will the new program require a network of repositories that provide long-term, permanent public access? Is the Lots of Copies Keep Stuff Safe model appropriate to consider to ensure redundancy in the new program, particularly in a state prone to significant natural disasters?

⁸² *LOCKSS for Government Documents: A Needs Assessment for the Federal Depository Library Community* (Stanford, CA: Stanford University Libraries, January 2004); available at <http://lockss-docs.stanford.edu/NAfinal.pdf>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

DSpace. DSpace was created to house research and publications for faculty and researchers at the Massachusetts Institute of Technology.⁸³ It is open source software built upon free tools such as Linux, Apache Web server and the Tomcat Servlet engine, and the PostgreSQL relational database system. DSpace offers the advantages of digital distribution and support for long-term preservation for a variety of formats, including text, audio, video, images, datasets, and more. It offers the opportunity to provide access to electronic information from multiple sources through one Web interface. Kansas ran a brief pilot test of DSpace, titled KSpace, to gauge its utility in a government documents environment to capture at-risk publications; however, large-scale implementation is currently being sought and no reports of the outcome of the pilot test are available.⁸⁴

NARA/SDSC. The work of the National Archives and Records Administration (NARA) and the San Diego Supercomputer Center (SDSC) has already been mentioned. Their collaborative work has included XML self-describing documents, storage resource broker technology, and data grid technology. NARA has also recently released a request for proposal and requirements documents for their Electronic Records Archive (ERA) project. The ERA development will bear watching, as it will be the first large-scale implementation of a digital archive for government information.

Open Archival Information System (OAIS) Reference Model. The OAIS is not a programmatic model, but a reference model that outlines the required functions of an archive. It was developed within the scientific community, but has great utility within the library and archival communities. It outlines the core functions of an archive in some detail, and has been used to build digital archive implementations.

RECOMMENDATIONS FOR ACTION

An OAIS-compliant digital archive should be implemented for government information with enduring historical value. The archive should be capable of receiving data from state publications repositories; and if necessary, sending digital resources out to repositories for access after preservation actions have been taken on them.

The functions of a digital archive will need to be developed and implemented as a separate functional unit, that works with the other components of the repository, and eventually, within the information infrastructure.

Section 3.6 Program Administration

Administering this new program will require a definition of roles and responsibilities by the stakeholder organizations, and an administrative structure that includes appropriate stakeholder representation and expertise. Collaboration is the one constant among all existing models and programs, as the distinction blurs between the traditional information professions and

⁸³ More information about DSpace is available at <http://www.dspace.org/>.

⁸⁴ "Kansas State Publications Archival Collection," *Crossroads: Developments in Electronic Records Management and Information Technology* 2003:2; available at http://www.nagara.org/crossroads/2003_2.html; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

technology. Managing and maintaining digital resources requires not just the expertise of libraries and archivists, but also of technologists. It requires both to work together if a satisfactory end result is to be achieved.

Sustaining coordinated action and effective communication among stakeholder organizations is perhaps the greatest challenge to be faced. The models for consideration below exemplify the work on these challenges in the library and archival communities. Note that the most mature and successful programs exhibited a number of common characteristics – legislation, partnerships and technical approaches.

FUNDAMENTAL QUESTIONS

- Which stakeholder organizations are appropriate for leadership roles and do they have the commitment and assets to take on leadership positions?
- What is the role and responsibility of each organization?
- What responsibilities will be shared and how?
- Are formal partnership agreements appropriate among stakeholders?

Models for Consideration

North Carolina. The Access to State Government Initiative was launched in 2002 in North Carolina and brings together a coalition of organizations “to conduct a three year project to research digital information issues, gain a better understanding of current publishing practices in state agencies, and develop solutions for managing state information in digital formats.”⁸⁵ The current initiative is led and staffed by the State Library of North Carolina. The effort is a collaboration between the State Data Center, the North Carolina Office of Archives and History, the North Carolina Vital Records Unit, and a core working group of primary stakeholders consisting of information providers, information facilitators and end-users.⁸⁶

Funding for the effort comes from a Library Services and Technology Act (LSTA) grant. In a white paper on the status of North Carolina digital state government information, coauthors Kristin Martin and Jan Reagan make a case that, due to their legal mandates to manage and preserve state publications and government records, the North Carolina effort is appropriately led by the State Library and State Archives.⁸⁷ This effort builds upon and brings new life into the FIND NC project begun in 1998 to improve access to government Web pages and information.⁸⁸

Preserving and Accessing Networked Documentary Resources of Australia (PANDORA). This effort is led by the National Library of Australia and is conducted through the efforts of nine partner organizations that contribute to a centralized repository and resource index. The National Library developed policy, procedures and a digital archiving system for building and managing an archive of Web publications and invited the state and territory libraries and other cultural collecting agencies to join it. The State Library of Victoria was the first partner to accept the

⁸⁵ Kristin Martin and Jan Reagan, op cit.

⁸⁶ Ibid.

⁸⁷ Ibid, p. 29-30

⁸⁸ FIND NC Web site, available at <http://www.findnc.org>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

invitation in 1998, and one by one, all of the mainland state libraries have joined, as well as ScreenSound Australia and the Australian War Memorial.⁸⁹

The PANDORA Web site identifies collaboration as vital to their program, stating, “The preservation of a collection of significant Australian Internet publications is beyond the capacity of any single collecting agency. A cooperative effort by the traditional deposit libraries, as well as other collecting institutions, is essential to achieve a broad and deep coverage of Australia's heritage published on the Internet.” PANDORA currently lists ten partner organizations.⁹⁰

U.S. Government Printing Office. In the United States, the Government Printing Office (GPO) is pursuing a number of pilot projects and partnerships designed to maximize the results of their efforts to update and modernize the Federal Depository Library Program (FDLP). FDLP is fundamentally a partnership program based on shared responsibility that is coordinated and led by the Superintendent of Documents.

T. C. Evans, Deputy Superintendent of Documents, outlined a number of pilot projects and partnerships currently underway in a speech at the 6th Annual State Government Information Locator Service Conference, April 2004, in Raleigh NC. He indicated that the GPO is constantly interacting with new groups, including private industry, to seek information and solutions that meet the strategic goals of the GPO.⁹¹

A key theoretical concept inherent within the GPO model is that of distinguishing between “*permanent public access*” and “*permanent preservation*” of the “record” copy. Like California, the federal information management infrastructure is segmented between GPO and the National Archives and Records Administration (NARA). The GPO is responsible for printing, distribution, and access to publications, and NARA is responsible for the selection and preservation of official agency records, including the subset of records that includes publications.

This distinction is unequivocally stated by the GPO, which claims that its electronic collection is *not* comprised of the record copies of digital resources, but of permanent access reference copies that are maintained by GPO or its partners. NARA is responsible for preservation of all agency records with enduring historical value. GPO further states that including a publication in its collection is not to be done in lieu of depositing the publication with NARA in accordance with records retention requirements.⁹² This model takes advantage of the expertise, workflows and capabilities of both entities.

The GPO is also working with the University of North Texas (UNT) libraries to provide permanent online access to electronic publications of selected former federal government

⁸⁹ National Library of Australia, *PANDORA Archive: Preserving and Accessing Networked Documentary Resources of Australia* [Web site]; available at <http://pandora.nla.gov.au/index.html>; Internet; accessed on 30 August 2004.

⁹⁰ The partner organizations are the State Library of Victoria, ScreenSound Australia, State Library of South Australia, State Library of New South Wales, State Library of Western Australia, Northern Territory Library & Information Service, State Library of Queensland, Australian War Memorial, Australian Institute of Aboriginal and Torres Strait Islander Studies.

⁹¹ GPO Access Web site, available at <http://www.gpoaccess.gov/about/index.html>; Internet; accessed on 30 August 2004.

⁹² U.S. Government Printing Office, *Managing the FDLP Electronic Collection: A Policy and Planning Document* (Washington, DC: U.S. Government Printing Office, October 1, 1998); available at http://www.access.gpo.gov/su_docs/fdlp/pubs/ecplan.html; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

entities. The collection, known as the *CyberCemetery*, includes federal sites that were previously hosted on *GPO Access* in addition to federal sites that have ceased operation.⁹³

RECOMMENDATIONS FOR ACTION

The State Library, State Archives, and other appropriate partners should meet as soon as possible to decide which organization(s) will be responsible for access to, and preservation and management of, state government publications in all formats. This is a key decision, as it will identify the organization that must take the lead in digital preservation efforts for California state government information.

Preferred Outcomes

Key stakeholders should meet and begin defining roles, responsibilities, and relationships. Key stakeholders should include the California State Library, California State Archives, CalRIM, Office of State Publishing, the state data centers, the depository libraries, the California Digital Library and the San Diego Supercomputer Center.

Stakeholders would discuss and commit to establishing an information infrastructure for California state government, plan for collaboration through committees and working groups, and generate memorandums of understanding identifying specific roles and responsibilities.

Preferred Model

All of the models described here are based on collaboration among various organizations with library, archival, records management, and technological expertise. Collaboration is crucial for the effort to establish a new program for preservation and persistent access to digital government information to be successful. During the March 2004 conference, the State Library asked for volunteers to serve on a steering committee to work on planning a new program, plus volunteers for associated working groups. This model is a valid way of bringing together groups with specific expertise to address particular issues and find solutions, with an umbrella committee to manage the effort, provide high level vision and decision making, and assure that working groups and technology test bed projects are within the scope of the overall effort.

Section 3.6a Roles and Responsibilities

Having reviewed the high level components of a revised depository program, a model for the roles and responsibilities begins to emerge. California is well positioned to move ahead with this initiative, with a variety of potential partners able to provide the necessary tools and expertise.

RECOMMENDATIONS FOR ACTION

Potential roles for the following agencies include:

⁹³ University of North Texas Libraries, *Cybercemetery* [Web site]; available at <http://govinfo.library.unt.edu>; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

California State Library. The State Library could:

- Serve in a leadership role in the overall development of the new repository program.
- Be a primary outreach organization to state agencies.
- Initiate the collaborative development of auto-notification and auto-distribution tools.
- Take a leadership role in collaboratively developing content selection criteria, methods and responsibilities.
- Coordinate activities and projects (working groups and projects must be led by the organization with subject expertise).
- Take a leadership role in issuing the requirements for repository management (the requirements document should be developed collaboratively), including metadata standards (these requirements, however, should be developed collaboratively).
- Serve in a leadership role in repository management.
- Act as the “repository of last resort” for publications that are not collected and managed by other repositories.

California State Archives. The State Archives could:

- Serve in a leadership role in the overall development of the new repository program.
- Be a primary outreach organization to state agencies.
- Take a leadership role in collaboratively developing content selection criteria, methods and responsibilities.
- Coordinate activities and projects (working groups and projects must be led by the organization with subject expertise).
- Take a leadership role in issuing the requirements for repository management (the requirements document should be developed collaboratively), including metadata standards (these requirements, however, should be developed collaboratively).
- Serve in a leadership role in repository management.
- Provide input and expertise into repository requirements, metadata standards, and data standards. It is vital that these requirements and standards be developed with long-term preservation in mind.

CalRIM. California Records and Information Management could:

- Serve in a leadership role in the overall development of the new repository program.
- Be a primary outreach organization to state agencies.
- Take a leadership role in collaboratively developing content selection criteria, methods and responsibilities.
- Coordinate activities and projects (working groups and projects must be led by the organization with subject expertise).
- Take a leadership role in issuing the requirements for repository management (the requirements document should be developed collaboratively), including metadata standards (these requirements, however, should be developed collaboratively).
- Serve in a leadership role in repository management.

Managing and Sustaining a State Government Publications Program in California

Content producers (State agencies). State agencies could assume several roles, depending on their level of interest and expertise. These roles may include:

- Serving as a repository.
- Providing content by submitting publications or notification of publications.
- Implementing metadata standards and data standards, as appropriate.

Depository libraries. Like state agencies, the depository libraries could assume various roles, again depending on their interest and expertise. These roles may include:

- Serving as a repository.
- Assuming responsibility for content selection of a specified group of materials.
- Providing access to the digital publications.

State data centers. The data centers could:

- Provide technical expertise, particularly in the areas of self-describing documents, data formats, and interoperability.
- Act as repositories.

California Digital Library. The California Digital Library has an ongoing interest in these issues. In particular, they could:

- Act as a repository.
- Provide harvesting technologies.
- Serve in a role analogous to the GPO/UNT cybercemetery.org project.
- Provide expertise.

San Diego Supercomputer Center. The San Diego Supercomputer Center has an ongoing research program investigating technology solutions to these issues. In particular it could:

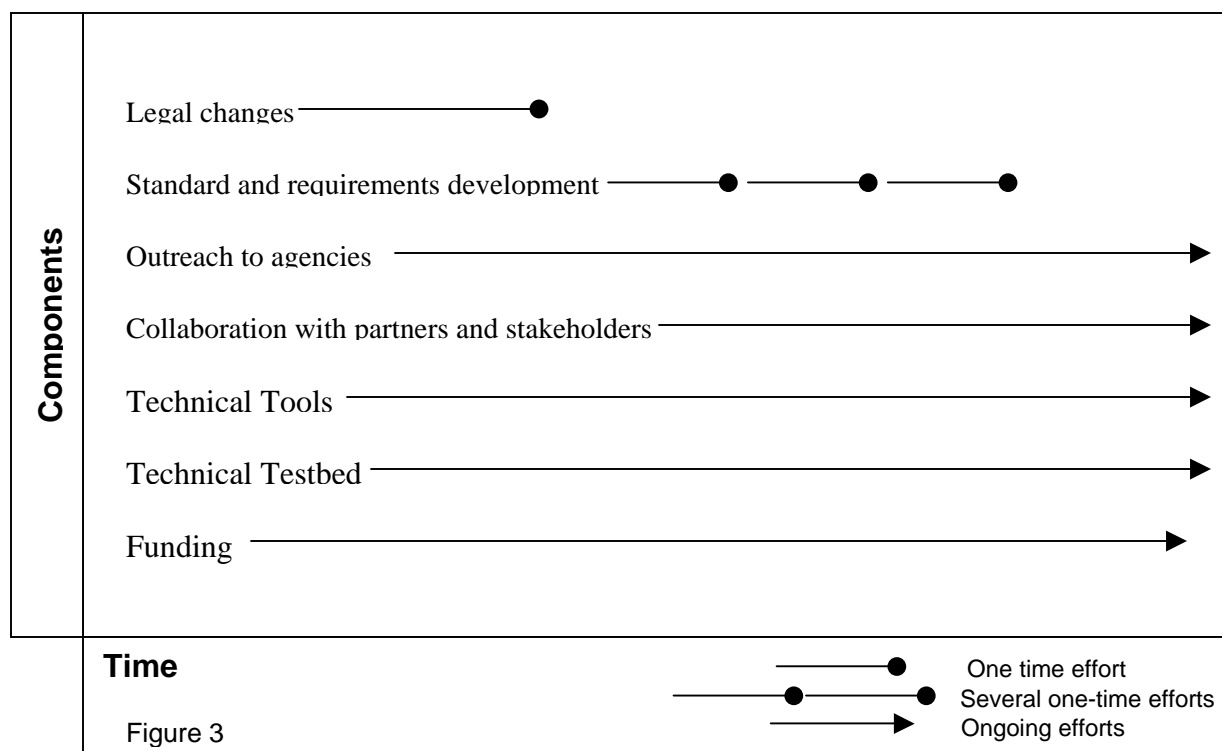
- Provide technical expertise, particularly in the areas of data grid technologies for management of and access to digital resources and technology standards.
- Act as a repository.

Section 4 Moving Forward

In order to move forward with this initiative, a coalition of many partners and stakeholders will need to be developed to build a new awareness of the need for a comprehensive and systematic program to collect, manage, preserve and sustain California government publications.

At the beginning of this report, we mentioned a digital government information architecture for California state government and the information infrastructure component that needs to be developed. These are necessary in order to effectively manage the information resources that are such an enormous asset to California government and citizens alike. In a perfect world, the infrastructure component would be developed before, and serve as the basis for and creation of the new repository program. In the real world, it is unlikely that this will occur. More likely, the new repository program will be developed first and serve as a model for the information infrastructure; or there may be parallel development on both. Even if no work is pursued on the infrastructure in the immediate future, the new repository program should be developed with consideration of its eventual place in an overall government information infrastructure.

Moving forward with the development of a new repository program will require significant work. Implementation will not be a linear process, but will more likely require several parallel initiatives that each work toward the common goal.



Managing and Sustaining a State Government Publications Program in California

Although the order in which these activities are initiated is somewhat dependent on opportunity, we suggest an implementation plan and strategy below. Further, we suggest that building a new depository program is analogous to building a structure or house. For a satisfactory outcome, there are foundational components that must be in place before other components should move forward. However, as the foundation is laid, one must be considering the next work to be done.

This implementation plan consists of three consecutive phases over a 9-15 year time line, as follows:

Phase 1: Focus on the framework (3-5 years)

- Identify program stakeholders
- Define stakeholder roles and responsibilities
- Initiate business planning
- Establish administrative structure

Steering/Oversight Committee

Purpose: High level visioning; political and budgetary support; evaluation of components to ensure they will move the effort forward.

Funding Working Group

Purpose: Identify and establish potential sources of one-time project funding and sustainable program funding. Implement a carefully selected Task Force to identify and establish sustainable program funding.

Legal Issues Working Group

Purpose: Identify appropriate legislative change. Implement specific task forces for particular areas of legislative work (i.e. copyright, definitions, incentives).

Standards Working Group

Purpose: Identify standards that need to be created, including metadata standards, data format standards, and repository requirements. Implement specific Task Forces for standards development and oversight of repositories.

Technology Working Group

Purpose: Identify technological components, identify and evaluate potential tools for implementation.

Content Working Group

Purpose: Create selection criteria (“must have,” “would like to have,” “would accept if offered,” “don’t want”); coordinate and promote depository selection based on criteria; ongoing evaluation of criteria.

- Identify and establish potential sources of sustainable and ongoing funding for the program (Funding Task Force)
- Enact appropriate legislative changes

Managing and Sustaining a State Government Publications Program in California

- Establish repository requirements
- Establish metadata standard
- Establish selection criteria based on macro-appraisal concepts
- Begin agency outreach initiatives to select agencies
- Begin discussion and implementation of “self-describing documents” (XML encoding) with select agencies that hold “must have” resources
- Initiate creation of repositories and access mechanism
- Establish technology test beds for several types of tools (manual harvest, bulk harvest, auto-notification, auto-distribution)

(It should be stressed here the program may have to rely on less-than-perfect technological tools. Bulk and manual harvesting will need to be considered at the outset in order to ensure that existing publications are not lost.)

Phase 2: Focus on implementation (3-5 years)

- Ongoing committee, working group and task force work
- Establish digital repository, ideally several, including access mechanism
- Coordinated outreach effort to “must have” agencies
- Auto-notification and auto-distribution operational
- Harvesting tools in place for select agencies
- “Self-describing documents” work continues with expanded list of agencies
- Ongoing evaluation and implementation of test bed technologies

Phase 3: Focus on evaluation and sustainability in a volatile environment (3-5 years)

- Ongoing committee, working group and task force work
- Established system for digital repositories
- Established access mechanism
- Operational digital archive
- Ongoing operation and evaluation of data capture mechanisms
- Ongoing evaluation and implementation of test bed technologies

Technology changes so quickly that it is difficult, if not impossible to anticipate what facets of the new program will entail into the future. The new program will have to adapt to new technologies, shifting political realities, and changing partners and collaborative options through time. In the initial years, the ultimate goal is to establish the framework (guidelines, requirements, workflows, legislation, etc.) that should remain stable and while acknowledging that the program will have to anticipate, expect and withstand change.

Funding and Sustainability

Like technology, government is constantly in flux. Changes in administration, legislators, political parties, agency organization, and available funding are all to be expected, and must be anticipated as the new program develops over time. While it is no revelation that accomplishing

Managing and Sustaining a State Government Publications Program in California

a redesign of the program will require people and money, there is little likelihood that either more staffing or program funding will be immediately forthcoming from the California legislature. The California depository program historically has suffered from lack of resources, with neither enough staff nor funding to conduct a viable program. It is clear that development of a new, effective program for access, preservation and management of digital state government information will require resources beyond those currently available.

The cost of developing a new program is a major concern. This was expressed clearly at the March 2004 Conference by participant David Harris, who wrote, “*I am also concerned that the State’s current imperative to do ‘more with less’ will lead to a focus on quick results which will not [be] satisfying in the long run.*”⁹⁴

There are immense resources, capabilities and expertise within California that are available to assist with the challenge of developing a new depository program. The most valuable resource is the willing participation of stakeholders throughout California government. The March 2004 Conference revealed interest from librarians and state agency stakeholder groups in joining this effort. Many participants left the Conference newly motivated to work on the problems, issues and potential solutions in developing a new program.

While the development of new tools and technologies to support the capture and distribution system is vital, people are indeed the most essential factor in such a reorganization and planning process. The most important aspect of a new program design will be the ability to collaborate, create partnerships and leverage the actions of stakeholders throughout California to contribute to the initiative. In the report, *It’s About Time: Research Challenges in Digital Archiving and Long-term Preservation*, Margaret Hedstrom notes that among the challenges framing the discussion of long-term preservation of digital information will be the ability of librarians, curators and archivists to re-design and automate current, labor intensive processes for management of analog and digital collections. Automated processes can be applied to the vast collections of analog materials as well as the ever-growing numbers of digital publications.⁹⁵

Successful consortia are able to build trust based on good communication and perceived benefit to all members. In the early, highly fragile phase of the new program’s development, good communication mechanisms will be essential. Building trust will depend on the ability of the leadership to maintain transparency for decision-making and consistent communication with all the stakeholder groups. The momentum generated by the March 2004 conference can be sustained by frequent and informative communication with the participants through e-mail, group lists and brief updates. Even for those participants who did not volunteer for the workgroups, having a sense of progress and inclusion in the process is important. Participant James Jacobs provided the following observation, “Avenues for continued participation must be

⁹⁴ See Appendix 1 of this report, p. 31.

⁹⁵ Margaret Hedstrom, *It’s About Time: Research Challenges in Digital Archiving and Long-Term Preservation: Final Report, Workshop on Research Challenges in Digital Archiving and Long-Term Preservation, April 12-13, 2002* (Washington, DC : Library of Congress, National Digital Information Infrastructure and Preservation Program, August 2003); available at http://www.digitalpreservation.gov/repor/NSF_LC_Final_Report.pdf; Internet; accessed 30 August 2004.

Managing and Sustaining a State Government Publications Program in California

mapped out soon after so that excitement level stays high...as we get farther and farther from the conference, that excitement dies off and other projects or work arise.”⁹⁶

Funding the new program needs to be placed in the context of business planning, and should be a key activity. The failure to act to identify and preserve important California state documents is costly. The fiscal consequences of failure to act should be documented and articulated where possible. Stakeholders should work with state agencies to identify such impacts and costs within their agencies. Members of the Funding Working Group need to be carefully selected and should be highly qualified to make the appropriate political connections. The members of this group will be the champions, lobbyists, spokespeople, and sales/marketing leaders of the new program’s development.

Funding options should include:

- The pursuit of grants for project funding. Grant funding will most likely be available for technology test bed projects, particularly testing new technologies and evaluating existing technologies.
- Participation in one or more projects with other states and the GPO. Arizona and Illinois are actively seeking partners to join with them in pursuit of realistic solutions to the same issues that California is facing.
- Consideration a fee-for-service cost model in partnership with the state data centers.
- Examination of model approaches such as the Washington State Digital Archives Investment Plan.
- Seeking operational funding from the stakeholder organizations to support activities that promote the new program.
- Development of a long-range funding request from stakeholder agencies to the California legislature based on specific pilot projects or capitol requests for equipment and technology. The most successful and long running state and national programs such as Washington, Georgia, Texas, Australia and the U.S. GPO are funded with public money. Illinois has been successful in obtaining national and state level grant funding along with public funding that supports staffing.

Conclusion

Establishing and funding a sustainable repository system to collect, manage and preserve state government publications, particularly digital publications, is a huge undertaking. Incorporating this program into a larger information infrastructure as part of a digital government information architecture is an even larger undertaking. The solutions involved must be creative and dynamic, yet based on law and standards. Core requirements for such an initiative must serve to create a flexible and sustainable program with committed partners and leadership. Technically, the program must adhere to standards-based and modular technological components in order to promote good management and sustainability of the digital resources. Ultimately, the goal is to create a program that consists of many interrelated projects.

⁹⁶ James Jacobs, University of California San Diego Libraries , E-mail to the March 2004 California State Library Government Documents Conference organizers, May 2004.

Managing and Sustaining a State Government Publications Program in California

California must develop new business models for the repository program that address key cost barriers, such as workflow and technology, intellectual property, and institutional costs and variables. The current high level of digital publishing activity by California government agencies requires that a state digital information management program provide structure, consistency, consolidation, and standardization to achieve realistic economies of scale.

The existing depository program and the program's stakeholders must create change that supports and recognizes cooperation in the California digital government environment. Environmental change will require sustained collaboration with a variety of partners, who must adopt a long-term view of the reworking of the depository program. Change may seem incremental at times; but will occur nonetheless. Positively, the new program will be developed in a state with a strong legislative history of promoting free public access to government information.

Finally, it should be clearly stated that California is not the only state addressing these problems; and California is not at all "behind" in addressing them. Both nations and states are looking for solutions that often do not yet exist. Yet significant progress in development of standards, procedures and technologies for providing preservation and persistent access to digital information has been made, and more will come. California can act now to ensure the long-term preservation of and access to state documents. A motivated and enthusiastic group of stakeholders already exists, the State Chief Information Officer is supportive of the effort, and the California Performance Review has specifically pointed out that the goals of the new program are vital for California. The time to act is now.

The people of this State do not yield their sovereignty to the agencies which serve them. The people, in delegating authority, do not give their public servants the right to decide what is good for the people to know and what is not good for them to know. The people insist on remaining informed so that they may retain control over the instruments they have created.--*California Government Code Section 54950*

Appendix 1

Final Report

California State Library

Government Documents

Conference

March 24-25, 2004
Sacramento, California

compiled by

Geri Bunker Ingram and Amy Lytle

Table of Contents

Introduction	2
Summary Report: Three Small-Group Breakout Sessions	3
Summary: Written evaluations as submitted post-conference	10
Transcription: Breakout Session 1—Legislation and Policy Issues	14
Transcription: Breakout 2—Practical Approaches to Working with Content Producers..	19
Transcription: Breakout Session 3—Building Sustainability: Technical Considerations .	23
Evaluation Responses: California Government Documents Conference.....	29
Follow-up investigation: California Government Documents Conference.....	41
Attachment 1	43
Pilot Project Sign-up Information: California Government Documents Conference.....	50

Introduction

On March 24-25, 2004 88 librarians, archivists, and state agency representatives, primarily from California, gathered together in Sacramento to address the issues surrounding digital government publications and their impact on the state's library depository program. The conference goals were to:

1. Educate participants about the issues surrounding digital government publications,
2. Spark discussion on solutions and approaches to the problems and issues raised by the speakers, and
3. Generate enthusiasm for subsequent actions, including the formation of collaborative working groups made up of people representing academic and public libraries, state agencies, and members of the state's archival community. These working groups will address various issues identified during the conference.

Participants had the opportunity to hear from a variety of speakers throughout the 1.5-day conference. Judith Cobb and Gayle Palmer set the stage by presenting a statement of the problem and an overview of the existing situation, both in California and in selected other states who are developing solutions to the problem of identifying, capturing, making accessible, and preserving digital government publications.

The balance of the conference was divided into three broad topic areas: legal, policy, and compliance issues; practical approaches to working with content producers; and sustainability issues, including sustainability in the technical arena (e.g., digital preservation), and in ensuring programmatic sustainability. Speakers presented various viewpoints and approaches to these issues. Each of the three topic area sessions was followed by facilitated small-group discussion. The participants were given a problem statement and three or four questions regarding the problem statement. The groups' answers and conclusions were shared at the end of each session.

These small-group discussions were the heart of the conference. The suggestions, identified challenges, and "brilliant ideas" that came out of the sessions provide insights that the California Government Documents Initiative can use as it moves forward. The documents that follow include analysis of the discussion groups' comments, post-conference evaluation comments, raw transcripts from the discussion groups, selected follow-up e-mail communications, and a table listing those who expressed interest in participating in various types of pilot projects for the Initiative. The conference participants' enthusiasm bodes well for California as they design a digital depository program.

Summary Report: Three Small-Group Breakout Sessions

At three separate times during the conference, the participants met in small groups to discuss the issues as presented in written problem statements and related questions. Each group chose both a recorder and a reporter for each session. The full transcription of recorders' and reporters' notes is found in this document.

Session #1: Legal, Policy, Compliance and Incentive Issues

PROBLEM STATEMENT

"State agencies are increasingly relying on electronic formats, including the Web, to make information available to the California public. Despite the widespread assumption that "everything is available on the Web," we know that the Web is not a reliable means of guaranteeing long-term access or information preservation. We need to find a way to ensure that state depository libraries are able to capture and store these documents, records, and other information for both access and preservation."

QUESTION 1: "Among the many state agencies, who are the key content producers?"

This question began a spirited conversation that endured throughout the 1.5-day conference. It seems every agency creates periodical reports as well as occasional publications—many are posted on agency websites and sometimes printed and sent to the various depository libraries. Each of the seven discussion tables produced an extensive list of agencies. These lists appear in the transcription in Appendix 1-C. The agencies represent elective offices, research and data centers, bureaus that provide services directly to the citizen, and industrial and regulatory agencies. Often repeated among the participants were various offices of the **legislature, environmental agencies, research bureaus and education offices**.

QUESTION 2: "Within the various agencies, who are the people/positions that should be included in any discussions about how to make sure state depository libraries are receiving the necessary information?" and "Who else has info or should be involved?"

Besides the agency representative who has the responsibility for producing the periodically released report, many publications come from a wide array of positions. The roles most often mentioned include webmasters, publication officers, administrative assistants/secretaries, librarians, records managers and archivists. With regard to who else should be involved in the publication process, many mentions were made of the **individual agency CIO** (chief information officer), the **Office of External/Public Affairs** and the **Legislature** itself. The **end-users** of the reports, whether in the form of citizen group or business leaders, were mentioned several times as stakeholders who must be consulted as the process of designing a digital repository for access and preservation moves forward.

QUESTION 3: “What are our incentives to make sure this happens—what’s in it for each of us?”

This question provided inspiration for reporters to make note of recurring themes, opportunities, challenges and “brilliant ideas”. (Again, full transcripts appear in the Appendices.) There was clear consensus that **reducing redundancy of effort** while **providing the basis for long-term public access** to unique government documentation comprise the most compelling incentives. Although it was recognized that **mandating legislation** is probably required to ensure programmatic success, alone it would not guarantee it, and legislative mandate should be pursued simultaneously with many other actions.

As the benefits of collaboration were further explored, it became clear that both the content-producing agencies and the depository libraries stand to gain in many ways. For the agency, training in metadata or even librarian-provided cataloging would be a big step forward in providing searchable, easily retrievable documents. With standard metadata applied, the documents could become part of a preserved repository under the library’s care. If the agencies proactively updated their regular publications, provided “purge dates” and kept their contact lists current, the librarian’s time would also be saved. The library could provide a one-stop shopping experience for the end user, and in doing-so link the public with its service agencies very transparently.

Emerging as the most cited keywords and phrases used in discussing these topics were **preservation after publication, standard metadata, and collaboration among institutions** (i.e., not only between individuals). When discussing challenges, opportunities and advantages, we often heard: **essential titles list, legislation** (mandating creation of a digital depository), **electronic documents retention schedules** and **cost**.

Perhaps the most compelling image was provided by the morning’s speaker Dean Mischynski’s (Director, California Research Bureau) idea of librarian as “**active participant**”. This suggestion captured the imagination of the groups as a whole, and evoked visions of active outreach, participation in the public’s education, “pushing” information and shunning the stereotypical role of “reactors”.

Session #2: Practical Approaches to Working with Content Producers

PROBLEM STATEMENT

“Not everything that’s published, regardless of format, is worth keeping. We need to come up with criteria for choosing what to keep.”

Two themes emerged quickly during the second breakout session. The first is regarding the need for a user-focused plan for a digital repository that is flexible (since the user groups are diverse) and not hindered by the traditional print-based depository model, where entities are either “selective” or “full”. The second theme surfaced as a somewhat tentative debate between two technical approaches to harvesting electronic data: “cherry-picking” or filtering only after a full harvest. It was evident that the participants enjoy varying degrees of interest and technological expertise, and they often expressed

opinions reflective of their subjective experience. E.g., some comments recorded included “*Free us to collect everything, now that it is technologically easy to do*” side-by-side with “*Collecting comprehensively now becomes completely unmanageable; we must have a systematic approach in place before we can collect comprehensively.*”

Underlying both themes is the need to fully explore the policies, principles and practices tacitly implied when speaking in terms of “decentralized” versus “centralized” services. After the conference concluded, many participants submitted well-considered evaluations that spoke to the need to devise secure (perhaps redundant) repositories and **centralized point-of-access interfaces** to end users, while providing **distributed functionality and responsibility** to both content-producers and to those maintaining the repositories. (See Appendix 1-B, Summary of the Evaluations.)

QUESTION 1: “How do we decide what’s worth keeping?”

Constituents of the participating agencies are obviously diverse; their needs must be better assessed. Aside from agency-mandated regularly released reports, there were few types of documents that seem obviously worth keeping to all participants. Attempts to identify general criteria reflected the uneven universe of possibilities as the groups named elements critical to *their own* primary clientele. Academic librarians were quick to identify as criteria **research value—e.g., statistical data and primary sources v. secondary materials**, while government agency providers, archivists and librarians noted **professional licenses** and **regulatory needs** in general as being indicative of importance for keeping.

Likewise, spirited debates ensued concerning balancing **historical value** and **uniqueness** with **current awareness**, especially with regard to information necessary for the maintenance of public policies.

The need for collaboration was not questioned, and several models for cooperation were sketched. One group cited the North Carolina model of creating “*Tiers of custodianship*” where materials were classed according to three levels of preservation need. Another group suggested creating hierarchies or schemes of cooperation to divvy up the functions of collecting and processing. Still another suggested that an advisory body be created among related institutions to oversee the appraisal process.

QUESTION 2: “Who should be involved in this decision-making process?”

The comments recorded during the discussions of questions 2, 3 and 4 were far less specific and focused than the opening question. Perhaps to be expected at this early stage in the conversation, individuals were somewhat hesitant to explicitly list the names or positions of agency personnel who need to be involved. Still the theme of collaboration was not questioned; partnering to achieve synergies seems a way of life among the California agencies, libraries and archives represented here.

It is clear that the conference participants value shared decision-making as well as the centrality of user-driven services. There were many comments reflecting a desire to get to know the agencies better, to survey their users and to **create working relationships with professional associations**.

QUESTION 3: “Where do we begin? (What’s our “low-hanging fruit”—and who has content to contribute immediately?)”

For those agencies for which the user-interest is central, the conversation turned naturally to which users need digital delivery most often. Following on the topic of diversity of user groups and needs, it was often stated that each group has a different culture and set of information-seeking needs and behaviors. One group reiterated the well-known query-negotiating paradigm: “[for each user group, we must] *distinguish between types of user needs: a. the need for a specific document and b. the need for whatever documents exist for a given subject or for a given producer.*”

Advocacy groups (such as environmental organizations), **historical societies**, and **community organizers** emerged as groups of particular importance in the discussion. One “brilliant idea” was simply to “*Poll the users*” to determine where to start.

For participants whose main responsibilities center more on the materials and their collection and preservation, the question evoked discussion of “model” state agencies with organized web sites, well-described and current materials. Brilliant ideas included “*Develop a systematic approach [that includes a] ...transparent/proactive system to acquire content from agencies.*”

QUESTION 4: “*How can depository libraries and content producers create effective partnerships?*”

Again, conference attendees are looking, by and large, to the State Library conveners for direction on building the coalition and proposing options for pilot projects. There is consensus that partnerships are critical for success, and that technological solutions created unilaterally will not ensure success. There is confidence that the public servants of California have the talent and the will to work together to create useful and dependable electronic document repositories.

Recurring themes from the session included **partnerships**, **different needs of different users**, and the notion that it is **people, not technologies, that solve problems**. Following on the user-centric theme, one brilliant idea was simply to **poll constituents** to learn their needs for a digital repository.

One concrete opportunity for working together that is not directly dependent upon a digital repository effort was that libraries should provide **union lists** to agencies, so that referrals are smoother. Another comment observed that **public policy making requires reliable and current statistical** and other **research**, and that libraries and agencies should work together to provide access to and to preserve these critical data sets in whatever formats they appear.

Session #3: Building Sustainability: Technical Considerations

PROBLEM STATEMENT

“Committing to preserving records in electronic format raises technical issues that are not encountered when preserving analog items. We need to define what we want our digital repository to “look like”—and to think about what our hybrid solutions might include.”

From the perspective gained during the previous sessions, participants were asked to set their imaginations free and to “blue-sky” about what they wished to create. Many of the same advantages and challenges discussed earlier re-emerged, perhaps more vividly colored. In order to put a framework around some of their ideas, they turned to existing models available in other states. This set the stage for volunteering to take up the work of follow-on task forces and working groups. Some of these volunteers will be asked to investigate the potential of joining in collaborative pilot projects, within California and with other states.

QUESTION 1: “What do we want our state digital repository to do?”

A very optimistic and yet realistic group of participants took time from their lunch to paint this picture of the functionality of the digital repository:

The State of California’s digital repository of government information will be easily searchable, with flexible interfaces customizable for varying constituencies. Results will be quickly attained, logically ordered for relevance, and easily synthesized. The full-text repository will not only be easily accessed, but will also provide standard cataloging (metadata) records that are authoritative, descriptive, and enduring as the repository’s technology undergoes inevitable changes.

Users and content providers alike will find materials in one place (regardless of where they are actually stored), and it will be easy to keep the material authentic, current, and relevant. Historical materials will also be easily found, and will be protected against inadvertent (or deliberate) removal or tampering.

The California State Library will act as a clearinghouse of catalog records, and may even undertake digitization projects in order to provide a coherent digital repository of core documents. There will be multiple copies of the repository stored in appropriate formats; media will be refreshed and platform migration through generations of technology will be assured.

QUESTION 2: “Who will use the digital repository and how will they use it?”

Conference attendees took a very broad view of the potential multiple audiences. Citizens, policy-makers, government employees, as well as state-run information technology management systems, will make use of the digital repository.

Equally diverse were the ideas for methods of use and access. Many specific ideas were put forth concerning advanced functionality and architecture for the systems. For end-users, it was reiterated that access must be transparent, intuitive, and

ADA-compliant. Since some users will still need mediation, while others will increasingly expect to serve themselves from locations remote from the libraries, both mediated and unmediated interfaces will be needed.

Although the question of whether and how the repository should or might be distributed was again raised, it was clear that centralized and multiple access points will be provided. Issues of architecture will need to be fully explored, tested, and measured for cost effectiveness and reliability during ensuing pilot projects collaboratively undertaken.

Because the participants themselves spanned a broad spectrum of interests and expertise, they offered a great deal of insight into how a digital repository (cf. analog and traditionally formatted information) could help to create new knowledge. For example, it was suggested that

"[one may] view documents, major and minor, as a line or series of related documents over time. Example: docs that support legislative histories."

Many references were made to geospatial information systems (GIS) technologies that can be integrated into a document repository that would allow for geographic mapping. Changes to the landscape, the environment, as well as policies and regulations, made over time can also be compared and measured through the use of digital technologies.

"...give geographic functionality to an environmental impact report (EIR) or County General Plan Repository."

Likewise, the cultural traditions that distinguish libraries from archives from research bureaus added to the rich mix of ideas concerning content and the need for understandable context. Some of the comments that triggered cross-cultural insights were:

"Legal documents are very sensitive—develop icons or legends to indicate, or [create] intermediate pages to explain..."

and

"Archival" state of content should be clearly stated—especially in dealing with historical versions of regulations."

Again the need to protect the private citizen while fully amplifying the public's right to information emerged in expressions such as:

"Authority of content is important." and *"Some data should be confidential (e.g., medical records.)"*

QUESTION 3: "Who might be involved in designing this repository?"

Stakeholders in the digital repository play roles such as cataloger, technology expert and standard advocate. Repository designers may include state archivists, CIOs, librarians at the State Library, the Office of the State Publisher, records and technology managers, as well as finance staff. Citizens and the state's university communities must also be involved at the design stage; public libraries will be especially needed to help with interface design.

The idea that the California State Library should lead the effort was expressed and supported many times. Equally emphasized was the value of the technological and scholarly expertise available and willingly offered from the State's many premier institutions of higher education. Models in play in other states such as Georgia, Washington, Illinois and Arizona were also invoked as being extremely valuable as aids

to design and implementation. Several speakers urged participation in collaborative pilot projects that are either already underway (e.g., Illinois: IMLS/CEP, UC/SDSC, LOCKSS). The success of the GPO Access system provided by the Government Printing Office underscored the potential and the need to create a manageable harvesting system.

During the final breakout session, many of the themes expressed in the earlier sessions were repeated with increased clarity, focus and caution. While it is true that the end-users need “*one-stop shopping*” it is equally true that “*different users need different interfaces*”. Concepts surrounding infrastructure crystallized into statements such as:

“Pro-centralization: concentration of resources [is needed] to fund preservation” but also

“Decentralized web organization enhances creativity and use to state agencies.”

Due respect was paid to the notion that standardization of metadata and technologies is critical to a sustainable repository. At the same time the participants recognized that a collaboratively built thesaurus and the creation of standard input templates would represent a tremendous commitment of resources in this diverse set of communities.

While obvious that momentum needs to be sustained, it was equally clear that funding and political realities in California indicate that a phased, inclusive approach is the only one that will be sustainable. Cautiously optimistic, participants suggested that California should not try to reinvent the wheel but should “leverage technology; implement the cool software architecture of the last three speakers [Cruse, Eckman and Marciano].”

A systematic plan needs to be developed that first ensures that all voices will be heard in the process of building a digital repository for the State of California. One “Brilliant Idea” stated simply: “*Funding models should be developed.*” Another was that next steps must include “*Recommendations to the legislature on what is needed by repositories.*” and “*Essential titles list necessary*” along with “*Notification system for new publications, pages, e-docs and **print** docs.*”

Policies and principles that were expressed during the conference need to be vetted widely and put to the test in efficient pilot projects. These projects should involve all sectors of the stakeholder communities, and should take advantage of efforts already underway in agencies within California. Attention should be paid as well to collaborative projects to which California has been invited to join. It is the collaborative atmosphere of the pilot test that will inform the feasibility of the technical, architectural and functionality suggestions made here in the convening conference.

Summary: Written evaluations as submitted post-conference

In order to facilitate the efforts of the State Library conveners as they move forward from the conference, the OCLC consultants prepared and distributed four questions in the form of a written evaluation. Of the 88 participants, 28 returned written evaluations, and twelve committed to further work through working groups, steering and advisory committees. All responses were extensive and thoughtful; an attempt is made here to summarize the ideas and concerns that hold sway among this group. A full compilation of responses is found in this report.

1. **In your opinion, what are the most achievable ideas generated from speaker comments and small-group discussion? What were you the most enthusiastic about?**

“The problem is bigger than one agency or library can do—we need to work together. I am enthusiastic about a plan that works with agencies (content producers) academic and public libraries (and CSL) to meet participants needs and benefit more citizens. This plan is vital to secure California State government information for future generations.” [Gloria Branson—CSU/Chico]

Participants by and large were confident that a **collaborative** effort, finding synergy among librarians, archivists, records managers and end-users, is not only necessary, but also enjoys a good prospect of success. This group by nature is committed to cooperation and distribution of labor, and **values the public good** highly. Methods of accomplishing common goals include joint task forces, diverse advisory and steering committees and full life-cycle planning of the envisioned repository.

Although there seems to be consensus that the end-user of electronic government documents should enjoy a **centralized access point**, nevertheless there is a lack of certainty as to which elements of the repository could be (and should be) distributed.

Elements that seemed to favor **distribution** include needs analysis and outreach/education. There is appreciation for the importance of end-user input into the access methods for the repository, as well as for providing a link between the end-user and the content creator.

“Educating the public and our suppliers is marketing by another name....and is exactly what is needed to make a dendritic solution work.” [Judy Weigel, LA County Public Library]

Likewise, materials selection and collection seem to be tasks that will require constant collaboration. *“Most achievable is probably to divide up areas to capture so that more libraries can help out. [Karen Aughinbaugh, Fresno County Free Library.]*

A sizable minority voice also speaks for distributing the task of document description (“cataloging” or metadata tagging) in order to achieve standardization. Given a choice between having the agencies trained to provide their own cataloging in an agreed-upon

standard fashion, and having the State Library provide the cataloging after submission or harvesting of the documents, there is no clear winner at this time. Some look to technology to provide automated ways to validate cataloging, through the use of controlled vocabularies and templates; some go so far as to look toward the development of “self-describing” documentation tools. Cataloging and metadata standards will surely be important areas for any emerging steering/advisory group to tackle early on.

Meanwhile, **centralization of the repository itself** seems to be very attractive, with well-articulated exhortations to keep multiple copies of whatever is deemed collectible. Whether the multiple copies are stored at disparate sites under different jurisdictions seems to be a matter for discussion. The LOCKSS and SDSC distributed “storage resource broker” methods were favorably viewed, though in general there was not a great deal of discussion of the technology platforms or architectural attributes.

Generally, all of the respondents were enthusiastic about what they experienced as inclusive involvement by the conference conveners, and most expressed high hopes for turning the initial momentum into tangible planning and knowledge-sharing sessions.

2. What concerns you the most about the process of developing a statewide permanent digital repository?

“Cost, as there is no way to do it cheaply; a program needs to be put in place. Technology is required and must be refreshed periodically. Selecting and securing information is labor intensive and making sure that it is adequately migrated so as not to be lost.” [Ben Amata, CSU/Sacramento]

Realistic and practical issues dominate the discussion where challenges are concerned. Overwhelmingly, there is acknowledgement that a shared solution will require extensive cooperation and that that in itself requires resources. In order to succeed, it is well recognized that there needs to be not only political buy-in but also active participation on the part of all content producers, libraries and archives, and that consensus-building must start with the early planning stages. Libraries and especially public libraries are seen as having the fewest resources to put toward the archiving effort. Meanwhile, one participant longed for traditional cost models:

“Archiving in a retrievable manner should be a part of the cost of the production of a government document. It was before (albeit in a procrustean fashion) and it should be now.” [Judy Weigel, LA County Public Library]

With diminishing budgets, participants are extremely worried that the cost of the effort, real or imagined, will work against sustainable success.

“I am also concerned that the State’s current imperative to do “more with less” will lead to a focus on quick results which will not [be] satisfying in the long run.” [David Harris, CERES]

3. What areas do you need more information about to make informed decisions about the directions a statewide permanent digital repository program should take?

Participants took this opportunity to underscore their concerns regarding the cost of undertaking a sustainable project. In particular, staffing issues were reiterated, with a desire to be “in the loop” but a hesitation to over-commit scarce staff. Again it was emphasized that although public libraries serve as depositories, some are at risk of elimination, and “*depository libraries are essential to democratic government.*” [Judy Weigel, LA County Public Library]

At least two respondents spoke to the need for a *method* of knowledge sharing among agencies, with one mention of Knowledge Management as a discipline and methodology for distributed organizations. Specifically, [David Harris, CERES] is interested in learning the cost to departments of *lost* information, and wants info on the state of the art automated classification systems. In order to better find a place within a collaborative undertaking, participants also seek knowledge regarding the relationships among, purposes and functions of each of the content providers. Another participant mentioned needing a “map of state government publishing” as a way to communicate quickly what is being published where.

Respondents also feel a need for more quantitative data about the problem. E.g., What percentage of the total number of State publications is NOT being printed in hardcopy? What would be the cost of participating in the various pilot projects that were presented?

Finally several respondents are looking for more information about the process itself. One wondered on paper: “*Are the State Library and Archives communicating “independently”* about the needs and the respective resources they might bring to bear? “*Is there already a game plan and if so what is it?*” asked one respondent. Several mentioned wanting to read the final report, as well as wanting to hear more specifics from the other breakout groups.

“Being in a group has the tendency to make that group’s discussion paramount, when in fact it either may reflect the thinking in other groups, or be totally off the wall as far as any consensus among the groups is concerned. Knowing what the predominant overall issues of concern by all the groups as a whole would provide a good foundation for developing informed decisions. [Ruth ?]

In any case, it would appear that all responding participants have an abiding and critical interest in being involved as stakeholders at every step along the way.

4. Are you interested in being a part of a working group to develop the repository program? Which topic(s) interest you?

There is overwhelming interest in participation in working groups. The “Sign Me Up” sheets that were circulated culminated in 34 volunteers across three focus areas: Steering Committee, Advisory Group and Pilot Projects.

Topics of interest include:

- Outreach / Education, to the public and among content providers and depositories
- Public Libraries (especially bringing the end-user stakeholder view)
- Steering, planning, policy-setting, (e.g., for deciding the distribution/centralization issues)
- Cross-organization (and across library types) collaboration
- Law and regulation
- System design, prototyping

There is also apparently a related interest group emerging from the Cal/EPA and they are already meeting to discuss these issues. The State Publishing Professionals, as they are calling themselves, were scheduled to meet on April 6th at the EDD in Sacramento. Aleta Zak reported that the group was open, and would like to continue a connection with the State Library.

In conclusion, the many participants who took the time to provide thoughtful evaluations are anxious to amplify their efforts by collaboratively forging an action plan and recruiting help from all stakeholders.

“If the intent of this conference was to send forth educated stakeholders to inform their constituencies, you’ve succeeded with me! – Judy Weigel, LA County Public Library

Transcription: Breakout Session 1—Legislation and Policy Issues

1. Among the many state agencies, who are the key content producers?

There are four sets of content-producing sites, of document types, of databases/systems of content.

- Legislature
 - Assembly Printing Office
 - Senate
 - Member web sites
 - Legislative analyst
- Governor's office
 - Finance
 - Demographic research
- Secretary of State
 - Voting
- Education
 - Frameworks for teaching
 - Test scores
 - Handbooks
- California Research Bureau
- Board of Equalization
- Office of the State Controller
- State Elective Offices—Attorney General, etc.
- Geological Survey (and other research agencies)
- Judicial Council
- Health and Human Services
- Energy Commission
- Food & Agriculture
- Environmental Protection Agency
- Coastal Commission
- Air Resource Board
- Fish & Game
- Cal Trans
- Equalization
- Justice
 - Crime Statistics
- Licensing/social services
- Health services in many agencies
- There are over 65 departments; any one may have its own IT department.
- DMV
- Franchise Tax Statistics
- Grants
- Water Resources
- EDD
 - Labor Market information (LMID a model?)

- Consumer Affairs
- Hoover Commission

2. Within the various agencies, who are the people/positions that should be included in any discussions about how to make sure state depository libraries are receiving the necessary information?

Players' Roles:

Public information officers
 Publication officers
 Technology officers
 Director
 Project managers
 Webmasters
 Report authors
 Librarians (or the equivalent within the agency)
 Admin assistants/secretaries
 Auditor
 Attorney
 Lobbyist
 Scholar
 Records manager
 Archivist
 Oversight

Who else has information or should be involved?

- Agency CIO'S--or techs or statisticians with databases might be the provider role, creating policy through practice? (E.g., CIO acts as lobbyist or enforcer for the public's right to information.)
- Freedom of Information Act (FOIA) (press releases, reporters make use of it).
- Office of External Affairs/ Office of Public Affairs.
- California Legislature
- Office of Planning & Research
- Leaders of citizens' groups—organizations, or businesses.
- Agency secretaries in the Governor's Cabinet are the ones enforcing policy.
- Users of all kinds are stakeholders.

3. What are our incentives to make sure this happens—what's in it for each of us?

- Payoff for agency security: backup and cataloging.
- Provide one-stop shopping for all the users.
- Improve online finding (a la Google).
- Save money with libraries taking care of storage.
- Library connects agency to citizens.
- We could give awards—superstar/most-improved agencies' sites.
- Public Records Act itself is an incentive.
- "It's our job to save this for you" (use good agencies as example to others).
- Reduce redundancy*
- Have the users of the information help develop description for key content.
- Metadata, tools provided by the library to the agency.
- Notification/push databases by the agency for the library.

- Reduce the workload (of each partner).
- Value-added by library to increase access.
- Make it easy for the agencies, e.g. give them a set of instructions.
- "Bind a set for them."

Recurring themes and keywords

- Permanence
- Quantity of publishing, agency reports
- Collaboration
- Preservation
- Education
- Legislative information
- Users
- Agency Functions
- Access by producer, by content, by item
- Outreach
- Standard Metadata
- Sites
- Databases
- Stakeholders
- Publishing policies and procedures
- Essential titles list
- Priorities
- Documents
- Not everyone collects everything.
- Divvy up the pie.
- To play a stewardship role means we need to talk together about who collects what.
- The activity is not so much "collection" as it is finding out what exists and getting it into a central depository.

Advantages & Observations

- Not limited to "publications" but all content
- Echo Dean Mischynski's (Director, California Research Bureau) idea of librarian as active participant.
- Find relationships among records managers, libraries, archives and state library.
- Often the most help comes from those in lower levels of the organization. **
- Sell the content producers on the value of their internal information and provide knowledge management assistance.
- Apply fees to state content modules to address retention and preservation.
- Cross-train librarians and database managers.
- Make the agency your customer.
- Look at who's keeping what to build collaboration.
- Rationalize "historic value" v. "current need".
- Synergies can mediate among differing values. Reciprocate services to create incentives for partnerships.
- Encourage dialog between model agencies and agencies that need help; the State Library could mediate.

- Electronic publishing pushing print costs to libraries and to end-users.

Concerns & Challenges

- Maybe agencies should also play a role in saving documents, mirror sets, archives in several places.
- ** You can't sustain this support (from the "grass roots" of the organization) without strategic direction from above.
- There are different levels of importance for accorded for retention and for access.
- Need collaboration, not reliance on individuals in agencies to enforce.
- There are laws, but there is no enforcement.
- There is no retention schedule for online material.
- Content producers often not in touch with the IT support who manage the website.
- What's "key" to one is not "key" to all. Selection criteria need to be established.
- Need to prioritize what's important to keep.
- Legislative mandates aren't always funded.
- Databases often show what is found, not necessarily what was asked for.
- It's often hard to identify, locate and communicate with proper parties (those responsible for publications' content).
- How do we regard the stuff that's received, but not cataloged?
- Agencies don't know libraries' needs.
- Define "publication".
- How do we measure the cost of retention?
- Need standardized metadata for electronic resources.
- What happens when old data disappears—need permanence after publishing.
- Existing documents guidelines are not really followed.
- Electronic documents are not easily available to the average user.
- We need snapshots of websites through time.
- Privacy issues: HIPPA makes search & retrieval more difficult; important content is frequently confidential.
- Records management guidelines and retention schedules are needed for electronic documents.
- Not all agencies' products are known to the state records management office.
- Records and archives/documents are related, but are fuzzy in the online environment.
- We can use the ca.gov portal to identify key producers, but we have a problem in not being able to crawl the databases.
- There is the problem of "cherry-picking" documents to preserve v. capturing whole the active websites and the ways in which they work.

Brilliant Ideas

- See ourselves as initiators, active participants, outreach, information *pushers* (not reactors).
- We need to preserve whole databases *and* their underlying structures.
- Sophisticated database management might be a role for libraries in the future.
- Might it be easier and cheaper to print out the original electronic file, scan it, and manage the derivative (rather than trying to archive the born-digital entity)?
- Push content changes as they occur.
- Provide a central site (state library) for e-distribution of all publications maintained.

- Bring public advocacy groups into this discussion/planning process.
- Identify the audiences.
- Cooperate to reduce redundancy.
- Use performance appraisals.
- Make creation of publications easy with easy to use tools and TRAINING.
- Recognize publicly who created the document.
- Think outside the box; if legislative bills are obvious to save, what about legislators' sites?
- Repurposing of data i.e. GIS for "recombinant potential"—should be a measure of what is "key" resource.
- Get business behind the effort.
- Move the legislation.
- Create MARC records to load into library catalogs.
- Use the extant models such as GALLILEO.
Document what you keep and what you discard, even if it's on the web.
- Every organization needs a CNO—Chief Nagging Officer!

Transcription: Breakout 2—Practical Approaches to Working with Content Producers

1. *How do we decide what's worth keeping?*

- Consider content worth separately from the methods that may be available to collect the content.
- Consider what we have lost, and reconsider how we might have kept it.
- Articulate: what is the inherent value in “old” things? In unique items?
- Consider what others are keeping, and collaborate to avoid redundancy. (I.e. consider keeping what is NOT being kept already.)
- Need to plan and to survey to find out what is critical. Contact agencies and their users.
- Keep agency-mandated reports.
- Do agencies themselves know what they want to keep?
- Agencies should keep a profile of what they select, and for alerts they send.
- Think of future constituencies too—what is ephemeral today may be critical in the future.
- Need to break away from current depository models based on print.
- Can collect so much more in the digital environment, but need to do some ranking.
- Collect all and assess later?
- Collecting comprehensively now becomes completely unmanageable; we must have a systematic approach in place before we can collect comprehensively.
- Tiers of custodianship:
 1. Collect it, leave it alone and store it.
 2. Create metadata to preserve it and make it accessible.
 3. Archive it: put in maximum effort for formatting, metadata, ability to back up, restore and migrate.
- Some depository libraries have no control over technology resources or are otherwise unable to make a commitment.
- If we store only easy-to-use materials such as PDF files, “snapshots” of websites, we can go only to 3 levels. This is helpful but we will still lose deeper levels. (Web snapshots can capture some functionality, but only at the level of web pages and PDF files. Not at the level of deep-web—cannot capture searchability of databases.)
- Different institutions need to keep different things; and need different levels of storage/cataloging access for same materials.
- Different constituencies need to be taken into account.
- Criteria for keeping:
 - Research value e.g., statistical data
 - Primary sources v. secondary materials
 - Professional licenses
- Regulatory needs. (Regulations themselves, but also, materials that reveal what the meaning and context is of regulations. i.e., supporting documentation)
- Hierarchy of repositories could collaborate to reduce redundancy among institutions.

- Separate functions of collecting and processing. One member would collect and another member institution would appraise, select and create metadata.
- Easier to do with electronic data than paper-based. Especially university system?
- Advisors from various related institutions would form an advisory or appraisal group.
- Full depository perspective: everything work keeping. But practically—can't get everything from day 1 –must prioritize.
- Many things thought useless end up being extremely valuable.
- Who are willing participants who can get us access immediately?
- When agencies decide to publish something they give it value and the need to collect. (E.g. water statistics.)
- Need things that are part of FOI requests such as agendas and local government body minutes e.g., County Law Library needs agenda and minutes; needs legislative history documents for decision-making and audit trails for public policy hearings.
- Selective depositories may collect the annual reports of the agencies themselves.
- Worth keeping:
 - Goals of each dept
 - Commission minutes
 - Summary of actions
 - Legislative hearings
 - Water documents
 - Unique and in demand items
 - Title 24 CCR
 - Old Bureau of Mines reports
 - Early 1900's bulletins
 - Cal state constitution

2. Who should be involved in this decision-making process?

- Agencies at different stages of evolution into electronic presentation/publication. Need to get to know and make contact with agencies with limited or undeveloped web presence.
- Suggest liaisons of local library association chapters, archival societies and historical societies with agencies to monitor and presser the work product.
- Professional associations, regional associations, e.g., ABAG etc. to provide clout and public relations and coalition.
- Users
- Faculty
- Content producers
- Collection development plan
- Government docs aren't correlated to academic departments

3. Where do we begin? (What's our "low-hanging fruit"—and who has content to contribute immediately?)

- Distinguish between types of user needs:
 - a. Need for a specific document
 - b. Need for whatever documents exist for a given subject or for a given producer.

- Advocacy groups: historical societies, environmental groups, communities of interest.
- Easy to get—based on format. (E.g. all PDF files are easy because they are discrete, static objects.)
- Keep track of reference requests—analyze request types and patterns.
- Look for collections that are already well-organized and thus lower-cost to capture, process, make accessible.
- Agencies with good websites.
- Address partnerships: who has and who wants.

4. How can depository libraries and content producers create effective partnerships?

- Essential to create partnerships—too much to do; can't afford redundancy.
- Purely technological solutions will not work. People getting together will solve the problems.

Recurring themes and keywords

- Different needs of different users
- Source of studies (legislative analyst office)
- Legislative intent
- Assess what you have lost to see what you would / should have kept.
- Primary v. secondary
- Partnerships
- Varied approach for specific and general needs
- Blue book
- User request
- Most often reason to archive: natural disaster
- Root documents
- Rights establishing ownership
- Collection development plans
- Do agencies know what's worth keeping?
- Education
- Build awareness
- Research value (statistics—databases)
- People solve problems

Advantages & observations

- Local libraries should collaborate with agencies to make documents freely accessible and provide a union list of who has what.
- Academia feels like a closed community, less a public-access agency.
- We are not always getting our government documents for free (e.g. geological survey).
- What's easy to get? What's best organized by the provider/producer (indexed, made available electronically).
- e-storage is not a problem—can collect more.
- Frees us to collect.
- How does all it "it" get collected?
- Does it get unmanageable?

Concerns & Challenges

- Fold in access to information (seamless access to all users).
- Regulations—agencies explanations for them.
- Uniqueness of info—unable to get elsewhere.
- Agencies—some have masses of info in Web-others have very little.
- Involve the agencies.
- Need systematic plan for collecting.
- Agencies need education and sharing, but there is no statewide communication among agencies, nor is there within the individual agencies concerning record retention. Need to change culture of agencies regarding info management.
- Agencies need to consider need for current as well as the need for historical record.
- Who keeps what?
- How to decide what's most valuable?
- Is it unique, old or hard to find or replace?
- In demand?
- What people need to do their job?
- Index/catalogued?
- Shelf listed?
- What are the users accessing?
- Find mandated/legislated reports.
- How do we know what will be important tomorrow?
- Should be done at state level or along with a university because it's beyond the resources of the local library.

Brilliant Ideas

- Divvy up expertise among different depository libraries—so one does not need to do it all.
- Poll constituents.
- Develop a systematic approach.
- Transparent/proactive system to acquire content from agencies.
- Professional associations: outreach/coalition-building.
- Identify specific NB collections—alert agencies to need for their input.
- Develop partnerships in specialty collections and storage/cataloging.
- Grab the server/
- First sort docs / items in terms of types rather by agencies/
- Government docs don't necessarily go to a single department within a university.

Transcription: Breakout Session 3—Building Sustainability: Technical Considerations

1. *What do we want our state digital repository to do?*

- Contribute and use descriptive standards with distributed data (like OCLC does).
- Provide redundancy for preservation actions.
- Provide full-text and catalog records that are authentic, well-described, and organized.
- Modeled on GPO Access, who will be the State equivalent of GPO? The State Publishing Office?
- What should we do about dynamic data?
- CSL, Archives, OSP, State records management (might become) “Cal Records management” (under a new name). I.e., these groups should collaborate.
- At the state level, content is not based on hard data as well as it should be, AND, we need to catalog it.
- At UCSD, when we come across a database, we catalog it. (We use OCLC to catalog, and Melvyl and SD Circuit as well as the CSL.)
- How would someone else know it is cataloged?
- CSL is a wonderful source—should act as clearinghouse of catalog records as well as digitizing projects.
- Blue sky: easily searchable, intuitive, search results, ranked by relevancy; retrieve things in small bunches. Vivissimo (search engine) has this concept.
- Provide information that we can use and in a way that is easy to use, fast, with lots of access/searchability.
- Illinois model good—could be a portal site for permanent public access.
- Look at Washington state model.
- Find content in one place, close to the agency that produced it.
- More interesting websites; well-designed and organized.
- State agency—have personal “spider” on the website.
- Repository—persistent presence of the electronic resource stays the same over years.
- Finding tool available too.
- Pull finding aid into own system.
- Watermark that guarantees authenticity.
- Transparent entry into local system.
- Central portal to access government data not have to know who is cataloging particular documents to find them
- Output:
 - Physically stored wherever
 - Choice of outputs: save it, zip it, print it, pull it into the catalog
- Citation—persistent location would mean not having to search again.
- Pull record into catalog and take them from local catalog to item.
- Use whatever the average citizen has access to: Google, Yahoo, next-generation Access.
- Public docs are “spiderable.”
- Manageable output, e.g., statistical abstracts and ADA compliant, voice output, make image on screen larger.

2. Who will use the digital repository, and how will they use it?

- Citizens, government information management systems.
- Back to Dean Mischynski's thoughts.
- Analogy to send recordings; increasing user expectations and you must meet them.
- Users stay the same, or may broaden.
- What do our end-users want? Today's end user v. future end users?
- End-users Google for everything—so what good are catalog records? (Refer to new partnership between Google and OCLC.)
- Coordination means cooperation.
- CSL can partner with other private institutions too; e.g., puts records in RLIN—we all use them.
- Can use Caldocs-L listserv as a communications tool.
- One question—what do you think is going to cause change?
- GIS component of State Parks—will put data in layers—spatial searching.
- Access needs to be transparent to users.
- Access needs to be centralized.
- Access needs to be ADA-compliant.
- Access needs to be unmediated.
- Internet or intranet?
- Links from agency websites to central access.
- Libraries want to use catalog records (in local catalog).
- Libraries could create local interfaces or web pages to help focus on heavily used sites.
- Multiple ways of access.
- Content and context are both important .
- Authority of content is important.
- Some data should be confidential (e.g. medical records).
- Should be good cross-references and links back to agency web site
- "Archival" state of content should be clearly stated—especially in dealing with historical versions of regulations.
- Legal documents are very sensitive—develop icons or legends to indicate, or intermediate page-explaining it. (But this needs to be very easy for customers to understand and not lose time.)
- Repository creates records and sends them to depositories with less funding.
- Ability to search by any type of attribute.
- Ability to convert docs from non-usable formats to current technology.
- Repository must have current and older format materials.
- Digitization of historical materials.
- Historical docs, materials using OCR.
- Verification process for authenticity. Example—California Roster is out of date; not now obvious on web.
- Experienced users—e.g., those that interface with government entities.
- Health users.
- Need information literacy.
- Benefits and entitlements.
- Consumer information.
- Business community information.

- Direct and public.
- Indirect and mediated.

3. Who might be involved in designing this repository?

- Many included in selection and acquisition from entire body, but there needs to be a defined whole.
- What is a depository library in this environment?
- Distributed LOCKSS concept is better.
- Central digital repository is then responsible, and this creates risk!
- State agencies also need the option to be a repository—this is vital with the deep web.
- Eye-readable as a last resort is mandatory,
- Do we need to keep everything digitally forever?
- Governor's office has overall control of the CSL website—but the digital repository will be different.
- CSL is supposed to get a copy of each document printed.
- Right now does not cost much to agency to put it up on web.
- How many agencies have their own library? Not too many.
- Roles: cataloging, technology experts, and standards advocates.
- Producers are most likely sources of cataloging/metadata—perhaps creating abstracts.
- Standards for agency names and for how to create metadata information—i.e., use drop-down menus to help creator fill out forms before entering creation of document.
- Important to capture metadata up front.
- Capturing, creating metadata should be done at time of publication, not creation.
- Notification of new data should be automatic (date-timestamp).
- Designers: state archivist, CIO, state library, office of state publisher, record management, IT structure, Department of General Services, finance people.
- Find business drivers and get them started.
- Libraries can be pathfinders—make needs known, do outreach) including to the legislature.
- All stakeholders—especially the Cal State Library, colleges and universities, public libraries, and the citizen off the street.
- CA State Library should lead.
- Colleges and university repositories.
- Public libraries interface with the public is a must.

Recurring themes and keywords

- One-stop shop.
- Transparency in searching (different levels, ADA compliance).
- Metadata for web (shallow and deep).
- Make metadata/cataloging part at the creation of the content.
- Different user interfaces.
- Redundancy = security.
- Is the depository concept applicable in the web environment?
- Longevity & redundancy = GPO model; born digital (captured).
- Can't happen all at once.

- Collaboration/participation.
- Capture everything.
- Pilot project.
- Definitions: document, draft.
- Coordination.
- Easy, intuitive, targeted search responses.
- Spatial plus topical searches made intuitive.
- Paper publication is still needed of records to be identified.
- Terabyte.
- Access points.
- Content v. context.
- Full depositories v. selectives.
- Selectives depend on full to have material may need in the future.
- If document is formally published, gives it a cachet for collectability.
- Adding items to the collection v. adding links in a catalog record.
- Ease of navigation.
- Full text search.
- Standard for web sites: GPO model. Centralized access point – distributed content.
- Persistence
- Transparency
- Public access

Advantages & observations

- Users need mediation in their searching.
- Phasing in implementation is reasonable.
- ADA compliance.
- THE solution is not completely apparent.
- Shouldn't lose sight of free access-- free for libraries and patrons.
- We have fewer face-to-face patrons and great patron demand for full-text online access.
- Real time access for producers on their intranets which can be moved to repository.
- Remote access but necessary connection to reference service to assist when necessary.
- Pro-centralization: concentration of resources to fund preservation.
- Decentralized web organization enhances creativity and use to state agencies.
- There is the California portal web that hasn't been explored or developed. Money is of course, involved.
- Sharing best practices.
- Most projects at data centers are supported by cost-recovery.
- Include procedures, depositories and control agencies (DGS, etc.).
- Two different models of data centers exist in the State. They need more coordination.
- Structures are changing as we speak.
- Consult users for need current & historical.
- Consult to learn what was sought and were they successful.
- Legal information.

- Full v. selective (bills v. regulations databases).
- Libraries have incentives to archive; what are the producers' incentives? (Help with metadata/indexing).

Concerns & Challenges

- Workflow for producer to user repository or user transactions.
- Funding
- Loss of information
- When do we think about legislation?
- Shared cataloging
- Collaborative thesaurus needed.
- So-called "minor" docs (agency meeting minutes). Are they more collectible now that they are published on the web?
- Major electronic document archiving—central authority doing it v. cooperative efforts of smaller libraries. Can smaller libraries collect a little outside their interests to contribute to the larger community?
- Diversity of requests.
- Need a way to make citation easier for users.
- Sites searchable by any search tools (spider, Google, etc.).
- Output be usable size and ADA compliance.
- Notify what's been discarded (paper or electronic).
- Capturing databases particularly for bursts of information.
- Happenstance of collecting.
- Survey of agencies' data.
- Uniform standards for formal publications.
- Risks of centralization of unique materials (consensus is "no").
- Agencies do not always use the same form of the agency's name.
- Separate content from application software is critical.
- Volatility
- Don't reinvent the wheel—expand on existing.
- Dynamic data: not really even now.
- Individual, uncontrollable state web access—spotty, etc.
- Duplication of effort.
- People-dependent.
- Transparent workflow.
- Templates from uniform standards.
- Need a thesaurus.
- Same models as in paper for elections.
- Organized process.
- Museum exhibit model.
- Depository searchable at the document level and at the "inside" the document level—metadata.
- Share cataloging information from a central authority.
- Need different user interfaces for diverse user communities.
- Authentication and trust security levels, validation.
- Distinction between archival and other copies.
- Funding for the infrastructure, e.g., problems with the California Portal for state agencies.

- Service for fee tradition for agencies (not for libraries).
- Universal involvement of repositories including local libraries; largest point of public contact.

Brilliant Ideas

- Viewing docs, major and minor, as a line or series of related docs over time.
Example: docs that support legislative histories.
- Personal spiders—respond to tag in agency sites.
- Central portal/directory so no need to know odd site locations.
- Data grid—choose from variety of output possibilities.
- Partner with Google—brand with icon to show that it is an authentic California government agency document.
- At public libraries, have a dedicated computer.
- User can take information away in many forms.
- Ask Ebsco and PAIS to include CA documents.
- Different access possibilities for different sensibilities.
- Role: collaborative pathfinders—value added.
- Leverage technology; Implement the cool software architecture of the last three speakers (Cruse, Eckman, Marciano).
- Automatic distribution of records or alerts (a la Marcive) to subscribers. From central library or agencies themselves.
- But who would deal w/this and how? Have a profile like the federal model?
- One stop shop.
- Give geographic functionality to an environmental impact report (EIR) or County General.
- Plan Repository.
- Guide to terminology (thesaurus), e.g., what does the Franchise Tax Board do?
- Link content back to producer.
- Reorganize agencies—4 groups: archive, library, records management, state publishing office.
- Clearinghouse of cataloging practices.
- Idea of core documents—need an essential titles list for California, similar to the one at FDLP.
- Panic Button
- Basic and Advanced searching or two versions of the site with different views of data for different levels of users.
- Next steps:
 - Funding models should be developed
 - Essential titles list necessary
 - Notification system for new publications, pages, e-docs and print docs
 - Recommendations to the legislature on what is needed by repositories
 - Ebsco and PAIS should be partners in indexing.

Evaluation Responses: California Government Documents Conference

1. **In your opinion, what are the most achievable ideas generated from speaker comments and small-group discussion? What were you the most enthusiastic about?**

One of the most achievable ideas from the event was to form a workgroup to pursue implementation. The level of expertise and commitment was exceptionally high.

(I'm most enthused about) the opportunity to put more of the non-GIS ideas developed at CERES into practice over a much broader subject matter. While it's far from trivial, I believe it is possible to develop distributed tools to support improvement of access to local information assets within departments which would use standard indexing, data structure and communications methods providing a local benefit which would provide an incentive to organize information assets so that they are accessible across all departments. – David Harris

Creating an immediate paper archive of all Web sites as the federal Government Printing Office is doing. – Aleta Zak

There will be a greater likelihood of success if all participate and move towards one goal. In the history of document retention, this is a landmark point in time that will significantly change the entire process. This conference significantly impacted my viewpoint for the need to preserve documents for generations to come. – Guy Blair

Evolving an active role for libraries in state government operations, the "dendritic" paradigm. Participating in a network of content producers. – Paul Veisze

The most achievable idea I heard discussed in our small group, which was echoed throughout the two days among many other participants, was a centralized "e-government" repository site that would be responsible for archiving government documents. I also like the idea of housing such an entity under the auspices of the State Library. I really like the idea of one website containing links to various agencies, archived documents, etc. that would serve as "one-stop shopping" for the citizens of our state. – Suzy Daveluy

I'm most enthusiastic about the awareness of the value of government information that can be created as we discuss these processes with both the content providers (i.e., state agencies, etc.) and with user communities (e.g., advocacy/issue groups, Sierra Club, etc.). I don't yet have a sense of what is most achievable. – Tom Bickley

I left the conference with the feeling that progress will be achieved! While I am not sure if concrete ideas stated or if a formal game plan was actually installed, I felt that there was a definitely consensus that change must take place and will

take place. It is not clear to me who the leader of this movement is (if there is a leader), or who will be driving the next steps. I felt all of the speakers were excellent, though I felt the timing of Liz Bishoff's piece was really poor. Her advice was practical and excellent, but she really put a damper on a conference that allowed us to think "blue sky" and to really imagine that we held the power to create change in the current depository system. It would have been nice to end the conference on a more positive note. – Natasha Kahn

Most achievable: a networked solution. In a strange way, having the State Library employees contact individual agencies, find their document gatekeepers and educate that gatekeeper to 'throw a copy over the wall,' gives value to individual effort. This initial personal effort must give way to an institutionalized one, but it would be good to maintain the working relationship. Continuing effort must be made to keep everyone on the same page. This conference was a great start. Marketing and PR aren't usually the purview of government agencies. Educating the public and our suppliers is marketing by another name. It seems to me this aspect has been neglected, and is exactly what is needed to make a dendritic solution work. – Judy Weigel

The thought that we don't have to do it all at once and that we need not be comprehensive. We can take small steps and keep building on that. I was very encouraged to hear Mr. Jewel say that the State Library is committed to doing something - however small it may be. Also, having a collaborative partnership whereby institutions agree to host part of the digital collection and the CSL act as a clearinghouse for these partnerships is achievable. – Sushila Selness

I feel that there needs to be one person or a group of persons in charge of the digitization, storage and electronic access of CA Gov Docs. I do not think the agencies should be in charge of deciding what gets put into electronic format and what doesn't. There should be a CA Gov Doc portal which is searchable by field (i.e. subject, keyword, author/agency, date, format, etc.) and there should be a thesaurus available. There need to be standards in place, such as there are for Gov Docs in print, for the Depositories to follow in terms of electronic information (access, storage, computers, etc.) and their role as a depository. – Karen Reitz

What I am enthusiastic about is the energy and enthusiasm of those involved in CA Gov Docs community. I am enthusiastic that a plan is starting to take place and that you are asking for the input of those who work with the public and must access/store the information every day. Personally, I am enthusiastic about Gov Docs in electronic format thereby eliminating the need for microfiche, unwanted print publications and physical storage space.

Developing some programmatic way of contacting agencies to keep information officers aware of the depository program so that information is constantly kept in the pipeline and fugitive or missing information is kept to a minimum. – Ben Amata

Amending the Library Distribution Act. Collaboration and cooperation at all levels – from creators to stewards to access providers. – Ruth Dye

Most achievable is the collaboration between agencies to find a solution to the problem. I was very enthused to see representatives from various state agencies attending--we need state agencies to buy into the digital access and preservation concept. – Gary Klockenga

To focus on a group of agencies and create a "test group" to implement new technology such as pushing materials electronically to either the CSL or the CSL and depository libraries. To create a PR program to entice agencies to cooperate through publicizing their materials and indirectly their mission. To create a link between CSL, a group of concerned citizens or associations who need this information, and depository libraries to publicize the need for agencies to comply. Begin setting goals and establishing standards for presenting the materials.

Focus on what we need and what we need to know (I presume this is what the report will tell us). I was very enthusiastic about the idea of pushing information. – Linda Muroi

Ability to hear from those who are already doing capturing and sharing with so many people from all types of libraries. Most achievable is probably to divide up areas to capture so that more libraries can help out. – Karen Aughinbaugh

The problem is bigger than one agency or library can do -- we need to work together. The California State Library needs to take a leadership role. I am enthusiastic about a plan that works with agencies (content producers) academic and public libraries (and CSL) to meet participants needs and benefit more citizens. This plan is vital to secure California state gov. information for future generations. – Gloria Branson

The conference attendees recognize that public relations and education efforts are needed to make Agency managers/senior staff more sensitive to the statewide need for document preservation. As difficult it is for the State Library to acquire Agency documentation, realize that even the Agencies have difficulty acquiring and managing on to their own documentation! A statewide education effort could help Agencies understand how to manage their internal documentation more effectively. I was most enthusiastic about the fact that the State Library invited State Archives to the conference. The two professions need to collaborate more... Inviting records managers to the conference is useful; records managers are a direct contact to our Agencies document creators; and could assist with supplying content to a statewide repository. Though, I'm not sure that the Librarians in attendance really understood the potential asset that Agency records managers represent to this endeavor. – Diane Voll

I definitely got the sense that real cooperation among State agencies, the CSL, and document librarians is achievable to address these issues. I'm most enthusiastic that we are going to work on plans and potential solutions before both the production of government information and the technology changes are both totally out of control. – Christine Gladish

I was most enthused about the group sense that we need to re-examine the whole concept of what documents are (and the depository program); also very pleased by the agencies' participation and enthusiasm. – Joan Allen-Hart

Collaborative/cooperative efforts – Dora Ng

Capture materials at risk; provide access as pilot project. Legislative

Re-definitions of key terms. – Linda Heatherly

Collaboration between producers, librarians, archivists, IT. Creation of a MARCHIVE type system. – Kathy Dabbour

Starting with a "test group" of agencies. – Dennis Hagen

Enthusiastic about tackling this problem in a distributed but uniform way. Achievable - educating myself on resources. Reporting on this process to other staff. Would like to see the California State Library approach agencies to collaborate. – Judy Weigel

Issues of collaboration – involvement of stakeholders – LOCKSS. – Sue Kendall

The sense that a competent, dedicated group of movers in the state government are looking seriously at the issues and are ready to take action. – Dennis Ladd

Any better understanding of the overall problem/need/issue. – Claudina Nevis

CSL communication and collaboration with libraries. – Yvonne Wilson

Digital archiving on the GPO model. The SALT research from the San Diego Supercomputer Center is encouraging in terms of making materials accessible in various computer formats. (No name)

2. What concerns you the most about the process of developing a statewide permanent digital repository?

I am concerned with how new this is. I didn't hear of anything comparable to what we need to achieve and the innovation required increases the risk a great deal. If the volume and specificity of material requires as much departmental cooperation as appears, I am concerned that the apparent level of cooperation required from the departments is very high and very uncertain. I am also concerned that the State's current imperative to do "more with less" will lead to a focus on quick results which will lead to the magnitude of the task being underestimated, resulting in a system which is not satisfying in the long run. I am concerned that the current interest in IT consolidation and existing turf may prevent use of a distributed architecture resulting in an architecture which may not be customized sufficiently for some departments causing them to withhold their essential support. – David Harris

The fact that the LDA does not cover electronic documents. It's hard enough to get agencies to comply with laws that do apply to them, but an all-voluntary effort might be impossible. – Aleta Zak

The concern I have for this project is the ability to get the funding needed to make the transition to a digital library. Two barriers will hinder the funding aspect: 1) The state's financial crises and 2) Educating the legislature and Governor's Office of the importance of document preservation. Their support is crucial. – Guy Blair

Will I be able to adequately represent my content? What will conformance cost me, specifically, migration of content from proprietary standards to open standards? – Paul Veisze

My greatest concern is getting the buy-in of all the state agencies, libraries, archives, etc. throughout the state. The exclusion of any of the major players will lessen the effectiveness of any product we design. In addition to the buy-in concern, I am also concerned about creating "input standards" that would work for all involved entities. While the idea of a "template" was at first somewhat appealing to me, I think it would probably be too simplistic for a state the size of California (i.e. with the large number of agencies/organizations that would be participating in this venture). However, I think uniformity is key in creating a database that is user-friendly and easily accessible, and therefore some uniform standards will need to be created. A related thought-- perhaps the uniformity might not be attainable from the input end of the process, but would rather need to be implemented from a "cataloging" perspective after the information is created. The desirable uniformity in this instance would be having the same agency doing the cataloging. – Suzy Daveluy

Development of procedures/incentives for content producers to proactively participate. – Tom Bickley

I am greatly concerned about costs (both in terms of \$\$ and staff time) to the Libraries. My library is a medium-sized Library with little extra funding and little enthusiasm for cataloging government materials. Hiring more staff or shifting staff responsibilities for cataloging or capturing electronic documents is not an option under the City's current budget situation or under the Library's current administration. – Natasha Kahn

The cost. Who is going to pay for this? Government agencies must be educated to realize that web publishing isn't any cheaper than traditional print. Electronic media just pushes the cost of printing and storage elsewhere. Archiving in a retrievable manner should be part of the cost of the production of a government document. It was before, (albeit in a procrustean fashion) and it should be now. My second biggest concern is directly related to cost. What kind of persistent permanent technology are we going to use to provide access? It will have to morph over time, and the additional readings make it clear that such efforts are underway. Sounds like lifelong learning for anyone providing access. It isn't how we can produce documents better with technology, it's doing things with those documents that couldn't have been done before. Hopefully that will include broadening access. – Judy Weigel

The first concern is for adequate resources - both funding and know-how. At the federal level we have GPOAccess as a model of what the permanent digital repository should resemble. GPO is a single entity controlling this but they too solicit partners and there are several good examples of partners in existence. At the state level this will be a critical thing - getting partners who manage subsets of the digital repository. During the conference, three entities stood out (CDL, UCSD and Stanford) as having the know-how to manage vast amounts of digital information. In a state the size of California, there need to be more such entities. I am concerned that there may not be enough other institutions that would take the burden off of the three mentioned above. – Sushila Selness

I am most concerned about access issues and costs for the libraries/patrons. I am a Serials/Gov Docs librarian in a relatively poor area (Inglewood, CA). In this area people do not own computers for the most part, nor are they computer/information literate. Our library does not have a great deal of money to afford computers, printers, CD-ROM drives, etc. Also, a MAJOR concern is that the public will be forgotten. We must remember who the end users are, they are not just academic researchers and government bodies, but they are the people of California, the public. In developing the process you need to pay attention to all the various users in the State. The academics will need access to reports going back many years, for instance. The public will need easy access to California Codes, Tax Information, Renters Rebate forms, etc. The plan needs to incorporate the needs of all users. It also needs to include all divisions of the CA State Government. For instance, the fact that the Agencies must PAY to post their information on the CA State Portal is ridiculous and will only discourage the agencies from providing information to the public. – Karen Reitz

Cost, as there is no way to do it cheaply; a program needs to be put in place. Technology is required and must be refreshed periodically. Selecting and securing information is labor intensive and making sure that is adequately migrated so as not to be lost. – Ben Amata

Commitment on the part of the participating institutes, agencies, etc. Regardless of the level at which these entities are involved, and that involvement will depend on the recommendations of the final report, commitment to carry out their assigned role is fundamental and essential. – Ruth Dye

Public libraries are the weakest link. We do not have the funds to acquire and develop the technology to achieve digital preservation. – Gary Klockenga

To figure out a way to pay for electronic storage. – Linda Muroi

Cost considerations and perhaps reinforcing the idea of many in Govt., as well as some in management in Libraries that everything is online and we don't need paper at all. – Karen Aughinbaugh

Communication with all parties and all levels. Knowing what others are working on so we don't duplicate effort or re-invent the wheel. Despite budget cutbacks we all face, we need to focus efforts to get best results w/little money. – Gloria Branson

The long-term obstacle is how does the State accomplish long-term preservation of electronic documents and their migration to new technology standards as needed. There are no standardized (for State government) IT tools, policy/procedures or budget to accomplish the tasks set out in the current Library Distribution Act. Also the current budget crisis in California will make a digital repository project a rather low priority on the political horizon. Perhaps California could look at the efforts of the National Library of Australia who have set up the "Digital Preservation Coalition" -- a forum for digital preservation. We need an organized effort like the Australian coalition to review and recommend tools, policy, and procedures for a statewide repository. – Diane Voll

The usual.... budgets, state bureaucracy, and conscientious political will. – Christine Gladish

That it may be viewed as this week's "hot topic;" that without ongoing "care & feeding" by the conference leaders, people will go back to the stress of their daily work and let their enthusiasm wane. – Joan Allen-Hart

Ability to sustain what we want to gather – will funding be continually available? Loss of different versions of documents. – Dora Ng

Need for selection criteria related to public policy goals that can be justified and defended as cost-effective. – Linda Heatherly

Time and funding and establishing standards. – Kathy Dabbour

Funding the physical repository. – Dennis Hagen

Funding. Electronic format doesn't change cost of documents – it just shifts the cost to the end user. – Judy Weigel

The problem is progressing faster than the solution. – Sue Kendall

What will be missed, and how easy it will be to use the great sea of data. – Dennis Ladd

The number of stakeholders. – Claudina Nevis

Random "keyword" searching to locate significant documents by anyone; choice of cataloging record documentation in OPACs. – Yvonne Wilson

The permanent digital depository must have a transparent interface but must be available through a single site. There is a need for detailed indexing. There must be planning for sufficient computer space and funding for maintenance over time and formats. There is a need for specific legislation. (No name)

3. What areas do you need more information about to make informed decisions about the directions a statewide permanent digital repository program should take?

What a good question! Some very large private sector organizations have already spent a great deal of time on Knowledge Management initiatives, and I believe they are essentially trying to solve the same problem. I would like to learn what they have done, and what they have found successful in very distributed organizations. I would like to have some in-depth information on the cost to departments of lost information so that this can be used to build a case for their participation in organizing their knowledge assets. I would like to know the state of the art in automated document annotation and cataloging. The cataloging workload implied by this effort would tend to be very labor intensive and the budget cuts have taken most organizations to extremely low staffing levels. The labor necessary to do this completely manually is likely to be unavailable in many organizations. If an automated system could suggest classifications it could increase speed and reduce the tedium in this type of effort. I need a more in-depth understanding of the system requirements of all the stakeholders involved. – David Harris

I would like to know what percentage of the total number of State publications are NOT being printed in hard copy, but are only posted online? Our agency posts all of our printed documents online in two formats, Word and/or Adobe PDF. All documents are also converted to HTML. We publish very few online-only documents. What policies are other agencies following? – Aleta Zak

The financial aspects - initial and ongoing costs. – Guy Blair

Obtaining administrative support for delivery on item 2. – Paul Veisze

Here I admit my technological naiveté. I'm not quite sure how the information in dynamic databases would be preserved. While I understand the "snapshot" idea, I don't understand how the "preserving agency" would be able to capture the database at just the right moments to ensure we are keeping the quality and quantity of information necessary. While I have a pretty good feel for the interrelationships among the participating entities at last week's conference, there is still a need (on my part only, perhaps) to understand the functions/purposes of each...especially as the working groups come together to plot a plan of action. If a project is realized that puts the preservation responsibilities on the State Library, then there will need to be a fair and equitable distribution of other responsibilities to other agencies to ensure that CSL has the staff and time to undertake the major work this project will certainly entail. Really understanding the strengths and primary roles of staff in the various agencies is vital to creating a really workable plan. – Suzy Daveluy

I'd like to see a "map" of state government "publishing." – Tom Bickley

I'm not too concerned about the technology. For one thing, I probably wouldn't understand the technical solutions. I just want them to be something I, or any member of the public, can use. The role of public libraries as depositories concerns me. In this time of continual budget shortfall, GS is ripe for elimination.

Maintaining a depository library collection takes space and an inordinate amount of labor. How can we focus our collection of government documents to better serve our public? How can we streamline the selection and processing? How can we afford to print out everything our patrons want? I think there is a place for public libraries as depositories. The format may have changed but the public's right to know has not. Depository libraries are essential to democratic government. – Judy Weigel

I would be interested in knowing if the Calif. State Library and the State Archives are communicating about these issues independently and what resources do they have to carry out the functions of a statewide coordinator (or clearinghouse) if a collaborative effort is launched. – Sushila Selness

I need to know what the State of CA has in mind in terms of Government Documents and access/storage issues for the future. I want to know if CA is considering making ALL Gov Docs electronic or only some. Will the Government make it easy for us as depositories to search and find information and will this information be all in one place? Is there already a game plan and if so what is it? – Karen Reitz

What technology can do to minimize the cost. – Ben Amata

I think I would like a synopsis of what was reported by the discussion groups as a whole at the conference. This would provide me with an overall picture of what was being discussed and emphasized by all the groups. Being in a group has the tendency to make that groups discussion paramount, when in fact it either may reflect the thinking in other groups, or be totally off the wall as far as any consensus among the groups is concerned. Knowing what the predominant overall issues of concern by all the groups as a whole would provide a good foundation for developing informed decisions. – Ruth Dye

We still don't have a good grasp on the costs associated with this, and we would assume them. – Gary Klockenga

The group list should work fine. – Karen Aughinbaugh

Redefining the California State Library depository program for complete and selective libraries. How do we get the California legislature to support this effort? – Gloria Branson

It would be useful to actively seek out the assistance of the States records managers and archivists: both professions have (or try to have) access to their Agencies published and non-published documentation. Libraries are looking for material that records managers and archivists may already be aware of. – Diane Voll

Better understanding of how creating metadata records will help bridge rapid and certain technological changes. – Christine Gladish

I would like the opportunity to read Gayle & Judy's report. – Joan Allen-Hart

Metadata, subject access. – Linda Heatherly

I have to think about that – a lot of info was presented. At least now I am starting to have an idea about what I don't know! – Kathy Dabbour

Want to be on an e-mail listserv, whatever, of what's happening with the repository program. – Judy Weigel

Space needs – any study predicting this? Cost, technical knowledge. – Sue Kendall

I just want to keep paying attention to the ongoing discussion and process. – Dennis Ladd

Security of a centralized server with long-term (at least 5 years) of financial support. Staff dedicated to repository not added on to existing jobs. – Yvonne Wilson

4. Are you interested in being a part of a working group to develop the repository program? Which topic(s) interest you?

I am interested in participating to the extent my duties allow. I am interested in all of it. My most effective contributions would likely be in the area of systems design, prototype development and liaison with departments within the Resources Agency. – David Harris

Possibly. It's up to my supervisor, Bill Albert, Printed and Web Publications Manager at CIWMB. I would need to see a list of topics, but I might possibly be able to raise awareness among other agencies. A few of us at Cal/EPA are in the process of starting a networking group of State Publishing Professionals (that's the name of the new group). Our next meeting is Tuesday, April 6, from 12 noon to 1 p.m. at the Employment Development Department, 722 Capitol Mall, Room 1063, Sacramento, CA 95815. (Please RSVP to me, Aleta Zak, (916) 341-6308, if anyone wants to attend.) We've already had one meeting and someone from the State Library showed up, so we would certainly like to continue the connection. I'm not sure how many other people involved in publishing for the State of California know about this problem, or if they are aware of the seriousness of the issues and their potentially damaging long-term effects. – Aleta Zak

Yes. I can assist with the outreach to other State agencies. – Guy Blair

Yes, very...esp. on content for digital geospatial data, AKA geographic information systems. – Paul Veisze

I am definitely interested in participating in a working group. The areas that interest me most would be the establishing of input standards and/or the development of a selection criteria plan (what gets saved, how often are databases "captured," etc.) I would hope that any and all working groups would reflect the diversity of the conference attendees, and to that end, I think I could bring the public library perspective to the table. – Suzy Daveluy

Yes! I'm interested in contact with the end user communities and the education/outreach aspect of the process. – Tom Bickley

I am interesting in being part of a work group. I would like to work to ensure that public libraries have a voice. – Natasha Kahn

How does a public library decide what is 'essential' to their patrons? How can we change the language of government documents to make them sexy? How can we change the language of government documents to more closely match what my patrons ask for? For example: They ask for the effect of television on children. We have:

Mass media hearings: a report to the National Commission on the Causes and Prevention of Violence Author: Baker, Robert K., 1948- Call Number: Pr 36.8:V 81/M 38/2 Publisher: Washington, D.C. : U.S. G.P.O., 1969.

The information is pertinent but the language doesn't come close to matching the requesting terminology. How can we streamline the selection process so I'm not getting a pig in a poke (item selection.)? How can we reduce the inordinate amount of clerical processing? How can I select stuff that will actually check out, or at least get some use? Thanks for inviting me to this conference. If the intent of this conference was to send forth educated stakeholders to inform their constituencies, you've succeeded with me! – Judy Weigel

I am interested and did sign up at the conference. I would be interested in being part of a liaison group that could contact a state agency and communicate with their public information officer regarding the need to preserve digital information at a central location (away from agency web site) to provide access to future generations. This is just one thought that comes to mind. I am open to being assigned other tasks. Being a law librarian I can represent the needs of law library patrons. We are always struggling to find agency decisions and orders from past years that seem to disappear when a new year starts. I would be interested in identifying (selecting) digital collections for preservation. I would also be interested in writing or editing guides or pathfinders if and when a sample digital collection goes up. – Sushila Selness

Yes. Public access issues. – Karen Reitz

Yes, the job of putting together a program is huge and will require input from the stakeholders and cooperation because ultimately the program will probably need to be distributed in nature. – Ben Amata

Not at this time, but I would like my institute to be kept informed. – Ruth Dye

Yes, I can help on a working group. However, we do not have funding for travel and related expenses, so I am limited to e-mail, telephone, FAX, etc. Since I am not well versed in some of the technology, I am better prepared to help with collection development/content capture or general policy issues. – Gary Klockenga

I am interested, but before I sign up, I would like to see the report that comes out in June and early directions to get a better idea where I can help. One area of interest to me is how a medium academic library (like CSU, Chico) can better serve remote users (including public libraries) covering large geographic areas. – Gloria Branson

I would be interested in a working group that would actively include the States records and Archives programs as a resource to access content producers. – Diane Voll

I'm planning to retire in a year or so, but I'm most interested in the need to have permanent, easy, access to the information. – Christine Gladish

Yes – as requested (maybe statutory/regulatory revision). – Joan Allen-Hart

No – have currently gone part-time and probably will be retired in about 1 year. – Dora Ng

Formation of legislative/regulatory program. – Linda Heatherly

Yes – assessment. – Kathy Dabbour

As a CSL staffer, I'm sure I'll be part of the process. – Dennis Hagen

Yes, particularly interesting in how an Internet based resource can serve non-internet literate users. – Judy Weigel

I am interested, but am not sure where I want to be on the topics. – Sue Kendall

Don't know – will continue to look, listen and consider what I can do. – Dennis Ladd

Already signed sheet – Yvonne Wilson

Follow-up investigation: California Government Documents Conference

On April 26, 2004, attachment 1 was written to the twelve individuals who had indicated that they wished to be involved in further work. Eight of the twelve recipients formally acknowledged the invitation to discuss pilot project activities. To date, two members, Paul Veisze, Research Manager 1 at the California Department of Fish & Game, and Lucy Barber, Archivist with the California State Archives, have replied with substantive comments.

The Fish & Game Department runs certification programs and a GIS system for many applications critical statewide. For example, they issue kelp harvest permits for the marine region; they also manage the State's Spill Prevention and Response unit.

Paul Veisze answered the email invitation both in writing² and by telephone. He included in his written response the names and contact information for the people with whom he is currently working but who were not present at the conference. (See NB, Attachment 1.) After listing three activities describing ongoing and new ways to collaborate with the State Library, Paul telephoned to explain his ideas and the projects more fully.

A. California Resources Agency California Advisory Committee on Geographic Names (CACGN):

Paul and Janet Coles have been working together on this important effort. Paul sees himself continuing as liaison and "matchmaker".

B. Fish Habitat Aerial Reconnaissance –Eel River and Tributaries, 1993:

Imagery and geo-spatial information. Again Janet has been involved with this collaboration and Humboldt State Library and the CSL are both interested in acquiring the work product.

C. The California Interagency Watershed Map of 1999 (Calwater 2.2.1):

This is a federal standard under review for watershed metadata. Other participants at the conference (and recipients of the follow-up email), Gary Darling and David Harris are partners in this CERES-based effort. Paul noted the websites mentioned in his email; perusing them together we discussed the phenomenal work being done by Robyn Myers. She is the main facilitator, web-wizard and chief "diplomat", doing archival work (and more) on this State-wide issue. In closing, Paul offered to send to the OCLC Consultants' group the current draft of the California Watershed metadata set. (This 42-page text document was forwarded for possible inclusion in the final report.)

In closing, Paul offered his thoughts on why the conference seemed to hit the mark, and why he is encouraged about continuing collaboration. He emphasized that, like the other

¹ Email correspondence, Geri Bunker Ingram, April 26, 2004, copy attached

² Email reply from Paul Veisze, April 26, 2004 copy attached

conference participants, he “is a practitioner”; involving State agency staff who are responsible for the daily operation of information systems at issue is fundamental to finding practical and sustainable solutions to sharing and preserving the government written record for the citizens of California.

Lucy Barber chose to send private email, the essence of which is reported here. Lucy credits the conference and especially the inclusion of Richard Marciano from the San Diego Supercomputer Center with triggering important new collaborations already. One of those efforts involves the SDSC, the California Archives, the State Library, and the Legislative Counsel in exploring joining a project in Minnesota dealing with electronic records and publications.

Lucy also reports significant progress “partly as a result of serving on the Advisory committee which gave me a better understanding of some practical issues”. Lucy is now fulfilling the duties of Electronic Records Archivist, duties that heretofore had not been fully examined. Using that title, she was able to arrange with the data center for the transfer of a number of websites “of agencies that have closed recently (including the former Governor's websites in a number of iterations)...”. She also is doing outreach to the Records Managers of the State and reports “...As a result of that outreach, I know have contacts in the Justice Department to follow-up with. Though my concentration is on "records" rather than publications, I expect these communications across agencies lines will also help the library in their efforts...”

It appears from Lucy's report that even the day-to-day communication between the State Library and the Archives has improved markedly since the conference brought principals together. She reports, “Other equally important advances have been made since the conference ...the conference created new *links between the state library and the state archives*, which means that the state archives has now revived a previously-moribund practice of transferring state publications that we do not want to the State Library....” Although not “fancy electronic web harvesting” this does advance the Library's efforts to acquire both historical and current documents. Further, this transfer is enabling the identification of some agencies which have *not* been transferring records to the library, allowing the Library to make the necessary contacts to rectify that situation.

Lucy Barber's follow-up activities include serving on the Advisory Board as well as involvement with the SDSC and Minnesota projects mentioned above.

Attachment 1

"Geri Bunker Ingram" <gbingram@adelphia.net> 4/26/2004 11:03:39 AM

Hello,

It was a pleasure to meet you at the California State Library Government Documents Conference last month. From your comments during the sessions and your messages to the consultants' group and the CSL, it's clear that the issues we began to discuss hold enduring meaning for you in your environment. We received many thoughtful written evaluations, and were gratified when twelve of the participants indicated a willingness to be further involved with by initiating a pilot project, or by incorporating these issues and stakeholders into an extant project in your organization.

With that in mind, I am writing to learn if, having had a little time for reflection and discussion in your institution, you would now like to:

1. contribute to the final report any further observations about the conference and the issues raised there (via email);
and/or
2. engage in a phone conversation where we might discuss what it would take to participate in follow-on activities.

Please feel free to reply to me directly, or to the group at large. (I'll append your names here so that you can see who comprises this group at present.) If you would like to talk further by phone, please send me a convenient time and phone number for you, and I will be happy to call you this week. Here are the sorts of things we'd discuss:

Are you currently engaged in a project or program relevant to the issues at hand that you would like to discuss?

Do you have an idea for a new collaborative pilot project?

If yes, what would be its focus?

Can you articulate some objectives that could be measured or evaluated?

Who would the partners be?

What would be your role?

Thanks very much for your continuing interest and enthusiasm.
Geri Ingram

Geri Bunker Ingram Consulting
www.gbingram.com
gbingram@adelphia.net
760.931.9313 (voice and fax)
760.525.8871 (mobile)

P.S. Here are the recipients of this group message. Certainly if I have missed someone who you think should be included, do pass this on to her/him, and please let me know their contact info:

Diane Voll
Sushila Selness
Ben Amata
James R. Jacobs
David Harris
James Scott
Gary Darling
Patricia Todesco
Kris Kasianovitz (by Christopher Coleman)
Yvonne Wilson
Paul Veisze
Lucy Barber

[NB: Added by reply copy from Paul Veisze and added here for reference—gbi]:

Robyn Myers	robyn.myers@ca.usda.gov
David Wagner	dwagner@conservation.ca.gov
Robert Sathram	rls@lib-mail.humboldt.edu
John Ellison	john.ellison@resources.ca.gov
John Quinn	jfquinn@ucdavis.edu
Robert Payne	rpayne@usgs.gov

Attachment 2

From: "Paul Veisze" <PVeisze@dfg.ca.gov>
To: <gbingram@adelphia.net>
Cc: <robyn.myers@ca.usda.gov>; <david@ceres.ca.gov>;
<dwagner@conservation.ca.gov>; <dwagner@consrv.ca.gov>; "Paul Veisze"
<PVeisze@dfg.ca.gov>; <kw1@humboldt.edu>; <rls@lib-mail.humboldt.edu>;
<jcoles@library.ca.gov>; <jjewell@library.ca.gov>;
<john.ellison@resources.ca.gov>; <jfquinn@ucdavis.edu>;
<rpayne@usgs.gov>; <gdarling@water.ca.gov>
Subject: Re: CA Documents Conference Follow-up
Date: Monday, April 26, 2004 1:57 PM

Hi Geri--

Thanks for an inspiring Conference. My reflections have evolved from initial, positive impressions of the Conference goals toward three lines of action. I'm sharing these after having discussed them with the people involved, ranging in time from pre-Conference planning to phone calls this morning.

As a GIS Manager within the California Resources Agency (CRA; Dept. Fish and Game), I envision the following areas of collaboration with the California State Library and Conference-initiated working groups:

1.) Liaison of CSL Government Publications (Depository Program) staff with that of the California Resources Agency California Advisory Committee on Geographic Names (CACGN): <http://cacgn.ca.gov> . See also U.S. Board on Geographic Names: <http://geonames.usgs.gov/bgn.html>

At the suggestion of CSL, a formal MOU would initiate this. I view significant, complimentary overlap in the standardization goals of depository programs and geographic names programs nationwide. (References: Roger Payne, USGS; John Ellison, CRA; John Jewell, CSL; Janet Coles, CSL; David Wagner, CACGN)

2.) Content focus (narrow scope): Fish Habitat Aerial Reconnaissance - Eel River and Tributaries, 1993: A unique integration of earth imagery and geospatial information. The California State University Humboldt Library and the CSL have expressed acquisition interest. Final documentation and rendering to DVD / CD media are in work, with release expected at the end of May. (References: Robert Sathram, CSUH; Janet Coles/CSL), and

3.) Catalog focus (broader scope): The California Interagency Watershed Map of 1999 (Calwater 2.2.1): A State-federal collaboration for standardized mapping, coding, and naming watersheds statewide: <http://www.ca.nrcs.usda.gov/features/calwater/> (interim host site). Future mirror to the California State GIS site, <http://www.gis.ca.gov> is under development. See National Biological Information Infrastructure, California node: <http://cain.nbii.gov> (References: John Ellison, CRA; Robyn Myers, USDA NRCS; Jim Quinn, UC Davis)

==

I'll follow-up with a phone call to discuss/schedule further roles and actions.

Thanks again for a great Conference. Looking forward to working with you and Team.

Paul

```
=====
Paul Veisze, GIS Research
California Department of Fish and Game
1807 13th Street, Suite 201
Sacramento, CA 95814
(916) 323-1667; Fax: -1431
pveisze@dfg.ca.gov
http://www.dfg.ca.gov
=====
```

From: "James R. Jacobs (Gov Docs)" <jrjacobs@library.ucsd.edu>
To: <gbingram@adelphia.net>
Sent: Monday, May 03, 2004 11:57 AM
Subject: Re: CA Documents Conference Follow-up

Hi Geri,

Sorry I didn't get back to you sooner, but I was on vacation for 3 weeks and am just now getting caught up with correspondence.

thoughts on the conference:

1. people get excited when they are asked to participate in an area near and dear to them.
2. preservation of state govt information is obviously an important issue, and one that involves more than just the state agencies producing the information. Libraries across the state need to be involved in a proactive role opposite of their traditional, reactive role as preservers of paper documents waiting for the state library to send them documents. New technologies make this both possible and imperative.
3. Avenues for continued participation must be mapped out soon after so that excitement level stays high. If you asked attendees at the end of the conference to do something, or give 1 hour/wk to a project, I bet 90% would've jumped on the chance. As we get farther and farther from the conference, that excitement dies off and other projects or work arise.
4. I'm not working on any projects at the moment. However, I have one in my pocket for which I'm hoping to write an IMLS grant with Richard Marciano. The project involves digitizing historic San Diego County planning documents. I have talked with Gary Klockenga at SD Public Library and Richard Marciano at SD Super Computer. Gary is interested and willing to give SDPL's copies of planning docs for digitization. Richard is interested in using Storage resource broker for storage and archiving of the digital files. He is also interested in using them in a UCSD class that he co-teaches on urban studies and for the researchers in the Regional Workbench Consortium (<http://regionalworkbench.org/>). I'm sure this project would help to inform any project involving the digitization of historic state documents.
5. I'd be excited to work on a pilot for distribution and preservation of CA state documents. Attached is the doodle that I made during the conference, which Richard showed during his presentation. I think this would make an excellent pilot, not only in setting up the infrastructure, but would give state depositories a chance to participate as well on the output end of the diagram. I'd be glad to talk with you and folks at CSL more about it.

Let's keep this discussion open. Could we set up a listserv for all the folks that you sent this email out to?

Regards,

James R. Jacobs
UC San Diego

James R. Jacobs
Government Information Librarian
Social Sciences and Humanities Library
9500 Gilman Dr 0175R
UC San Diego
La Jolla, CA 92093-0175
(858) 534-1266
<http://govinfo.ucsd.edu/>

Democracy is like blowing your nose. You may not
do it well, but it's something you ought to do
yourself

--G.K. Chesterton

"A library is an arsenal of liberty." Anonymous

(\n
{|||8-
(/

James R. Jacobs, UCSD Libraries

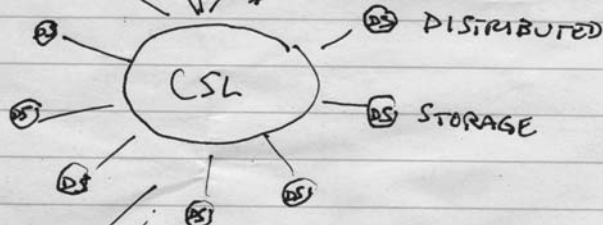
A GEN CIES



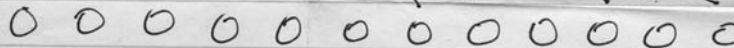
NEEDS:

- STATUTORY FIAT
- BUDGET
- COLLABORATIVE PARTNERS TO CATALOG, ARCHIVE...
- METADATA STANDARDS
- STORAGE
- FORMAT STANDARDS
- USER INTERFACE
- INDEXING
- DESIGNATION OF CSL AS RIGHTS HOLDER, OFFICIAL COPY,

RSS TO PUSH "DOCUMENTS" INTO CSL WORKFLOW



~~RSS TO PUSH CATALOG RECORDS TO LOCAL CATALOGS~~



CA STATE DEPOSITORIES

SELECT DOCUMENTS OF INTEREST FOR INCLUSION IN LOCAL CATALOGS

Pilot Project Sign-up Information: California Government Documents Conference

Interest Expressed	Name	Institution	E-mail	Phone	Notes
Steering Committee	Jacquelin Siegel	CA State Library, Information Technology Bureau	jsiegel@library.ca.gov	916.653.3441	Either I or another person from the CSL Information Technology Bureau would like to be included
	Natasha Kahn	Pasadena Public Library	nkahn@cityofpasadena.net	626.744.4279	?
	Patsy Inouye	UC Davis	pcinouye@ucdavis.edu	530.752.5247	Complete Depository for CA rep?
	James Scott	Sacramento Public Library (Central)	jscott@saclibrary.org	916.264.2795	
	Janet Martorana	UC Santa Barbara	Martoran@library.ucsb.edu	805.893.8724	
	David Harris	CERES	David@ceres.ca.gov	916.653.8092	Coordination of departments; technical assistance as needed
	Christopher Coleman	UCLA	cdgc@library.ucla.edu	310.825.0559	Preservation intent
	Gary Darling	CA Dept. of Water Resources	gdarling@water.ca.gov	916.653.5601	
	Susana Hinojosa	UC Berkeley	shinojos@library.berkeley.edu	570.653.9347	
	Ann Latta	Stanford Univ.	annlatta@stanford.edu	650.725.1131	Interests: Organization Workflow Levels of Access – Heterogeneity

Interest Expressed	Name	Institution	E-mail	Phone	Notes
Steering Committee (Con't)	Lucy Barber	CA State Archives	lbarber@ss.ca.gov	916.651.9419	
	James R. Jacobs	UC San Diego	irjacobs@ucsd.edu	858.534.1266	
	Ben Amata	CSU Sacramento	Bamata@csus.edu	916.278.5672	
	Yvonne Wilson	UC Irvine	ymwilson@uci.edu	949.824.7362	
	Bill Behnk	Legislative Data Center	Bill.behnkc@lc.ca.gov	916.445.4966 x 7795	
Agency Participation in Pilot	James Scott	Sacramento Public Library (Central)	jscott@saclibrary.org	916.264.2795	
	David Harris	CERES	David@ceres.ca.gov	916.653.8092	Coordination of departments; technical assistance as needed
	Christopher Coleman	UCLA	cdgc@library.ucla.edu	310.825.0559	If UCLA participates, Kris Kasianovitz would be the best choice
	Paul Veisze	CA Resources Ag. Dept. – Fish & Game	pveisze@dfg.ca.gov	916.323.1667	
	Gary Darling	CA Dept. of Water Resources	gdarling@water.ca.gov	916.653.5601	
	Patricia Todesco	CA State Archives	ptodesco@ss.ca.gov	916.653.2385	
	Lucy Barber	CA State Archives	lbarber@ss.ca.gov	916.651.9419	
	Diane Voll	CA State Parks	Dvoll@parks.ca.gov	916.324.9852	
	Sushila Selness	USD, Legal Research Center	selness@sandiego.edu	619.260.6813	

Interest Expressed	Name	Institution	E-mail	Phone	Notes
Agency Participation in Pilot	James R. Jacobs	UC San Diego	iriacobs@ucsd.edu	858.534.1266	
	Ben Amata	CSU Sacramento	Bamata@csus.edu	916.278.5672	
	Yvonne Wilson	UC Irvine	ymwilson@uci.edu	949.824.7362	
Other (serving on advisory group, etc.)	Natasha Kahn	Pasadena Public Library	nkahn@cityofpasadena.net	626.744.4279	?
	Patsy Inouye	UC Davis	pcinouye@ucdavis.edu	530.752.5247	I would like to serve here also
	James Scott	Sacramento Public Library (Central)	iscott@saclibrary.org	916.264.2795	
	Janet Martorana	UC Santa Barbara	Martoran@library.ucsb.edu	805.893.8724	
	Mark Pothier (Web Content Manager)	Judicial Council/Admin Office of the Court	Mark.pothier@jud.ca.gov	415.865.7448	Advisory Group
	Anthony Anderson	Univ of Southern CA	anthonya@usc.edu	213.740.1190	Advisory Group
	Arvetta M. Downs	CA Legislature Assembly	Arvetta.Downs@asm.ca.gov	916.319.3997 or 916.319.2846	Advisory Group
	Carol Doyle	Fresno State	Carol_doyle@csufresno.edu	559.278.3033	
	Gary Darling	CA Dept. of Water Resources	gdarling@water.ca.gov	916.653.5601	
	Dennis Ladd	Southwestern School of Law	dladd@swlaw.edu	310.826.0407	
	Alice Youmans	UC Berkeley Law Library	ayoumans@law.berkeley.edu	510.642.0296	

Interest Expressed	Name	Institution	E-mail	Phone	Notes
Other (serving on advisory group, etc.)	Susana Hinojosa	UC Berkeley	shinojos@library.berkeley.edu	570.653.9347	
	Joan Allen-Hart	San Diego Co. Pub. Law Library	Jallen-hart@sdcppl.org	760.940.4351	
	Susan Kendall	San Jose State Univ.	Susan.kendall@sjsu.edu	408.808.2039	
	Kristin Martin	State Library of NC	kmartin@library.dcr.state.nc.us	919.807.7445	If needed for preservation/access
	Sheila Bruton	CA Dept. of Education	sbruton@cde.ca.gov	916.445.0736	Dept./Agency Info
	Christine Gladish	CSU Los Angeles	cgladish@calstatela.edu	323.343.2015	Advisory Group or search engine or end user
	Tom Bickley	CSU Hayward	tbickley@csuhayward.edu	510.885.7554	End user education/publicity
	Therese Cason	San Francisco Public Library	tcason@sfpl.org	415.557.4472	Selection criteria advisory committee
	Ruth Schooley	The Claremont Colleges	Ruth.Schooley@libraries.claremont.edu	909.607.7122	Can participate in advisory group – State Docs & Statistics
	Diane Voll	CA State Parks	Dvoll@parks.ca.gov	916.324.9852	
	Sushila Selness	USD, Legal Research Center	selness@sandiego.edu	619.260.6813	
	James R. Jacobs	UC San Diego	irjacobs@ucsd.edu	858.534.1266	
	Lucia Orlando	UC Santa Cruz	luciao@ucsc.edu	831.459.1279	Advisory Group or other opportunity
	Bill Behnk	Legislative Data Center	Bill.behnk@lc.ca.gov	916.445.4966 x 7795	Also others

Appendix 2

Additional Resources: States and Copyright

States and Copyright – Setting the Context for Discussion

The National Research Council's study, [The Digital Dilemma](http://www.nap.edu/html/digital_dilemma/), contains an excellent chapter on "Public Access to the Intellectual, Cultural, and Social Record." It provides a general discussion of the issues for federal and state government information. A point of interest is that even at the Library of Congress the current system of copyright deposit needs modification to enable preservation. The publication is online at http://www.nap.edu/html/digital_dilemma/.

Jules Larivière's *Guidelines for Legal Deposit Legislation* is a revised, enlarged and updated edition of the 1981 publication by Dr. Jean Lunn, IFLA Committee on Cataloguing. Chapter 6, "Legal Deposit of Electronic Publications," provides an interesting case for establishing a "legal deposit" status for an organization. Such status would confer an official right to collect materials and preserve materials to maintain the social and cultural heritage of a nation. It also discusses the legal challenges relating to copyright. The publication can be found online at <http://www.ifla.org/VII/s1/gnl/legaldep1.htm>.

Additional background information can be found in the following publications:

Vaver, David. "Copyright and the State in Canada and the United States." Osgoode Hall Law School York University, Toronto, June 1995. Available online at <http://www.lexum.umontreal.ca/conf/dac/en/vaver/vaver.html> (accessed July 28, 2004).

Although this was written for a Canadian audience, Vaver provides a comparison of Canadian and US policy regarding copyright for government publications, along with suggestions for policy changes.

Samuelson, Pamela. "Digital Information, Digital Networks, and The Public Domain." Draft paper for a conference on public domain, Duke University Law School, November 2001. Available online at <http://www.law.duke.edu/pd/papers/samuelson.pdf> (accessed July 28, 2004).

This draft conference paper examines the implications of recent legislation, including UCITA, CIIA, and DMCA/SSSCA on the digital public domain. Although the focus is not on government publications, this provides helpful background on the legislation and its impact.

"GC Forum on Access to Government Contract Works." In *FLICC Newsletter*, Fall 2003, p. 3-4. Available online at <http://www.loc.gov/flicc/pubs/fn0303.pdf> (accessed July 28, 2004).

This article outlines a three-part series of discussions held by the Federal Library and Information Center Committee (FLICC) General Counsel Forum on copyright in government works. The series focuses on intellectual property rights as applied to works done by government contractors.

Examples of States' Policy Language:

Several states have good models for policy language. In addition, LexisNexis offers the State Capital product (<http://www.lexisnexis.com/academic/1univ/stcap/default.asp>), which gives subscribers access to bills, statutes and constitutions, digests of state law, proposed and final regulations, legislative directories, and news. Although the service covers all fifty states, some types of information are not available for all states. For example, proposed regulations for the state of Colorado are available online, but the final regulations are not available because of copyright restrictions.

Links to representative examples of copyright policy language are listed below.

Indiana: The state portal, *Access Indiana*, has a good “terms of use” policy found at: <http://www.in.gov/ai/policies/termsfuse.html>.

Maine: Maine has developed policies covering copyrightable works created by state employees; see <http://www.maine.gov/ispb/POLICY/COPYRIGHTABLE%20WORKS%20POLICY.htm>.

Michigan: As in Maine, works created by state employees belong to the state. Michigan’s policy addresses royalty payments. See Michigan Department of Civil Service Regulation 1.05 Patents, Inventions, and Copyrights: http://www.michigan.gov/documents/Reg1-05_20738_7.pdf

Mississippi: As in Indiana, the state web site, *Mississippi.gov*, offers terms and conditions of use: http://www.mississippi.gov/terms_&_conditions.jsp

Nevada: The state of Nevada has developed policy for intellectual property compliance; see State of Nevada e-Government Steering Committee Policy Control No. 3.02 Rev. Title A, “Web Style Guide,” [http://www.nitoc.nevada.gov/NITOCdocs/3.02\(A\)_Policy_WebStyleGuide.pdf](http://www.nitoc.nevada.gov/NITOCdocs/3.02(A)_Policy_WebStyleGuide.pdf)

South Dakota: State legislative language for contracts for use of state’s copyright: <http://legis.state.sd.us/statutes/index.cfm?FuseAction=DisplayStatute&FindType=Statute&txtStatute=2-16-8>

California: California state agencies frequently post copyright disclaimers. Sample language is found on the following web sites:

<http://www.ciwmb.ca.gov/Copyright.htm> - General
http://www.parks.ca.gov/default.asp?page_id=1021
<http://www.post.ca.gov/Chouse/pubs/copyright.htm>
http://www.consrv.ca.gov/CGS/disc_copyright.htm

The first link above, to the California Integrated Waste Management Board, provides an example of comprehensive copyright information and terms of use.

The California Lieutenant Governor's Web site addresses intellectual property issues at <http://www.ltg.ca.gov/programs/caip.asp>. This page discusses the lack of a comprehensive policy for state intellectual property and the application of federal law.

The Bureau of State Audits (BSA) prepared a comprehensive report on the issue of intellectual property for the legislature, titled *State-Owned Intellectual Property: Opportunities Exist for the State to Improve Administration of Its Copyrights, Trademarks, Patents, and Trade Secrets*. This was written from the standpoint of economic gain, but it gives a good picture of the state of affairs. It did not address the state university systems, which have their own sets of intellectual property rules. The BSA report can be found at: <http://www.bsa.ca.gov/bsa/pdfs/2000-110.pdf>; a follow-up report is located at <http://www.bsa.ca.gov/bsa-htmls/pdfs/sr2002/2000-110.pdf>. According to the follow-up, as of the time of its writing in 2002 no action had yet been taken on the recommendations of the original report.

International:

For a good exploration of the issue of copyright and access, visit the *Copyright and Licensing for Digital Preservation Home Page: CLDP Project*, Loughborough University (United Kingdom). The aim of the CLDP (Copyright and Licensing for Digital Preservation) project is to investigate whether copyright legislation and licensed access to digital content threaten the ability of libraries to provide long-term access to that content and to suggest ways in which the problems can be overcome. This Web page is located at <http://www.lboro.ac.uk/departments/lis/disresearch/CLDP/index.htm>.

Appendix 3

Bibliography

Bibliography

- Alberta Government Services (Canada). *Information Assets in the Government of Alberta: A Management Framework*. Edmonton, Alberta: Alberta Government Services, October 2003.
<http://www.im.gov.ab.ca/imf/pdf/IMFrameworkReport.pdf> (accessed 30 August 2004)
- Alemneh, Daniel Gelaw, et al. "A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience," *First Monday* 7:8 (August 2002).
http://firstmonday.org/issues/issue7_8/alemneh/index.html (accessed 30 August 2004)
- Allen, Nancy. "Collaboration through the Colorado Digitization Project," *First Monday* 5:6 (June 2000).
http://www.firstmonday.org/issues/issue5_6/allen/index.html (accessed 30 August 2004)
- Allen-Hart, Joan. Speech, California State Library Government Documents Conference, Sacramento, CA, 23 March 2004.
- American Library Association. Federal Documents Task Force on Permanent Public Access. *Federal Documents Task Force On Permanent Public Access: Final Report*. Chicago, IL: Government Documents Roundtable, American Library Association, December 29, 2003.
http://tiger.uic.edu/~aquinn/access/final_report.pdf (accessed 30 August 2004)
- Archives New Zealand. *Electronic Records: A Vision and Policy for the New Zealand Government Sector*. Auckland: Archives New Zealand, November 2002.
<http://www.archives.govt.nz/continuum/dls/pdfs/g1-electronic-recordkeeping.pdf> (accessed 30 August 2004)
- Archives New Zealand. *Electronic Records: Where Are We?* Auckland: Archives New Zealand, 2002-2003. No longer available on the Web as of 30 August 2004.
- Arizona. *Arizona Revised Statutes, Section 41-1346*.
<http://www.dlapr.lib.az.us/records/1-gov.cfm> (accessed 30 August 2004).
- Arms, Caroline R. "Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress." *RLG DigiNews* 4:3 (June 15, 2000).
<http://www.rlg.org/preserv/diginews/feature1> (accessed 30 August 2004)
- Australian Digital Theses Project: ADT Program Information Page. Web site. Sydney: University of New South Wales Library, February 2, 2000.
<http://www.library.unsw.edu.au/thesis/adt-ADT/info/info.html> (accessed 30 August 2004)
- Austrian National Library. *AOLA - Austrian On-Line Archive*. Web site. Vienna: Austrian National Library and the Department of Software Technology of the Vienna University of Technology.
<http://www.ifs.tuwien.ac.at/~aola> (accessed 30 August 2004)
- Authenticity in a Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.
<http://www.clir.org/pubs/abstract/pub92abst.html> (accessed 30 August 2004)
- Barata, Kimberley. *Archives In The Digital Age: A Study For Resource*. London, UK: Missenden Consulting for the Museum Library and Archives Council, November 2002.
<http://www.mla.gov.uk/documents/id435rep.doc> (accessed 30 August 2004)
- Barker, Joe. "Finding Information On The Internet: A Tutorial. Invisible Web: What It Is, Why It Exists, How To Find It, And Its Inherent Ambiguity." Berkeley, CA: UC Berkeley, Teaching Library Internet Workshops, January 2004.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html> (accessed 30 August 2004)

Barnum, George D. and Steven Kerchoff. "The Federal Depository Library Program Electronic Collection: Preserving a Tradition of Access to United States Government Information." In *Preservation 2000: An International Conference on the Preservation and Long Term Accessibility of Digital Materials*, 7/8 December 2000, York, England. Mountain View, CA: Research Libraries Group, 2000.

<http://www.rlg.org/events/pres-2000/barnum.html> (accessed 30 August 2004)

Barton, Mary R. and Julie Harford Walker. "Building a Business Plan for DSpace, MIT Libraries' Digital Institutional Repository," *Journal of Digital Information*, 4:2 Article No. 173 (May 7, 2003).

<http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Barton> (accessed 30 August 2004)

Beagrie, Neil. *National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity*. Washington, DC: The Council on Library and Information Resources and the Library of Congress, April 2003.

<http://www.clir.org/pubs/reports/pub116/contents.html> (accessed 30 August 2004)

Biehl, Kathy. "The Internet Archives: Preserving the History of Web Pages," *LLRX.com* [electronic journal]. Published November 1, 2001.

<http://www.llrx.com/columns/webcritic12.htm> (accessed 30 August 2004)

Blair, Guy (Administration Manager, California Office of State Publishing). Interview, 5 December 2003.

British Library. *Collection Care: Digital Preservation*. Web site. London, U.K.: British Library.

<http://www.bl.uk/about/collectioncare/digpresintro.html> (accessed 30 August 2004)

Building a National Strategy for Preservation Issues in Digital Media Archiving. Washington, DC: National Digital Information Infrastructure and Preservation Program, Library of Congress, April 2002.

<http://www.clir.org/pubs/reports/pub106/contents.html> (accessed 30 August 2004)

CAL Facts: California's Economy and Budget in Perspective. Sacramento, CA: Legislative Analyst's Office, December 2002.

http://www.lao.ca.gov/2002/cal_facts/econ.html (accessed 30 August 2004)

California. *Government Code. Section 14900-14912* [Library Distribution Act.]. Sacramento, CA: State of California.

<http://www.leginfo.ca.gov/calaw.html> (accessed 30 August 2004)

_____. *State Administrative Manual. Chapter 3100, Publications and Documents*. Sacramento, CA: California Dept. of General Services.

<http://sam.dgs.ca.gov/TOC/3100/default.htm> (accessed 30 August 2004)

_____. *Style Implementation Guidelines, HTML guide*. Sacramento: Teale Data Center.

<http://www.webmasters.ca.gov/styleguide/META> (accessed 30 August 2004)

_____. *Welcome to California: Conditions of Use: Ownership*. Web site. Sacramento: State of California.

http://www.ca.gov/state/portal/myca_homepage.jsp [scroll to bottom of page and click on "Conditions of Use"] (accessed 30 August 2004)

California. Governor (Schwarzenegger). *Executive order S-13-04 by the Governor of the State of California*. Sacramento, CA: Office of the Governor, Aug. 24, 2004.

<http://www.governor.ca.gov> [click on "Press Room on left menu bar, then on "Executive Orders] (accessed 30 August 2004)

California. State Chief Information Officer. *California Information Technology Council*. Web site.
<http://www.cio.ca.gov/ITCouncil/default.html> (accessed 30 August 2004)

California Performance Review. "State Digital Records Vanishing." In *A Government of the People for A Change: Report of the California Performance Review*. Sacramento, CA: California Performance Review, August 2004, vol. 4, section GG 45.
<http://www.report.cpr.ca.gov/cprprt/issrec/gg/eff/gg45.htm> (accessed 30 August 2004).

California State Auditor. Bureau of State Audits. *State Owned Intellectual Property: Opportunities Exist for the State to Improve Administration of Its Copyrights, Trademarks, Patents and Trade Secrets*. Sacramento: California State Auditor, Bureau of State Audits, 2000.
<http://www.bsa.ca.gov/bsa/pdfs/2000-110.pdf> (accessed 30 August 2004)

California State Library. "Enhancing Access to Electronic Resources Through the CSL Catalog: A Vision Statement." 2003 internal publication.

California state publications: manual for acquisition, processing, use. Sacramento, CA: California State Library, Government Publications Section, 1961.

CAMiLeon: Creative Archiving at Michigan & Leeds: Emulating the Old on the New. Web site. Ann Arbor, MI: University of Michigan, 2003.
<http://www.si.umich.edu/CAMILEON> (accessed 30 August 2004)

Cedars Project: Curl Exemplars In Digital Archives. Web site. Leeds, UK: University of Leeds, 2002.
<http://www.leeds.ac.uk/cedars> (accessed 30 August 2004)

Chapman, Stephen. "Counting the Costs of Digital Preservation: Is Repository Storage Affordable?," *Journal of Digital Information*, 4:2 Article No. 178, (May 7, 2003).
<http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Chapman/chapman-final.pdf> (accessed 30 August 2004)

Chief Information Officers Council (U.S.). Interagency Committee on Government Information. Web Content Standards Working Group. *Draft Recommendations on Web Content Policies to be Submitted to the Interagency Committee on Government Information (ICGI) by the Web Content Standards Working Group*. Washington, DC?: Chief Information Officers Council, Interagency Committee on Government Information, April 9, 2004.
http://www.cio.gov/documents/ICGI/Web_Content_Policies_DRAFT_4_9_04.doc (accessed 30 August 2004)

Cho, Junghoo and Hector Garcia-Molina. "The Evolution of the Web and Implications for an Incremental Crawler." In *Marking the Millenium: Proceedings of the 26th International Conference on Very Large Data Bases, Cairo, Egypt, 10-14 September 2000*. San Francisco: Morgan Kaufmann, 2000, p. 200-209.
<http://www.vldb.org/conf/2000/P200.pdf> (accessed 30 August 2004)

Coles, Janet (California Documents Librarian, California State Library). Speech, 6th Annual State-GILS Conference. 31 March 2004.

_____. Written comments, November 2003.

Collecting and Preserving Open-Access Materials on the Web. Web site. Ithaca, NY: Cornell University.
<http://www.cs.cornell.edu/wya/LC-web> (accessed 30 August 2004)

Colorado. *Colorado Library Law. Part 2, section, 24-90-204, Deposits of state publication*. Denver: Colorado Dept. of Education.
<http://www.cde.state.co.us/cdelib/LibraryLaw/Part2.htm> (accessed 30 August 2004)

Colorado Digitization Program. *Colorado Digitization Program*. Web site. Denver, CO: Colorado State Library.

<http://www.cdpheritage.org> (accessed 30 August 2004)

_____. *General Guidelines for Descriptive Metadata Creation & Entry*. Denver, CO: Colorado State Library, Spring 1999.

http://www.cdpheritage.org/resource/metadata/std_metadata.html (accessed 30 August 2004)

Communications Canada. Depository Services Remodeling Committee. *Proposal For A Revised Model Depository System*. Ottawa, ON: Depository Services Remodeling Committee, Depository Service Program, Communications Canada, June 2, 1998.

<http://dsp-psd.communication.gc.ca/Rapports/Monty/monty-e.html> (accessed 30 August 2004)

Cruse, Patricia (Director, Digital Preservation Program, California Digital Library). Interview, May 2004.

Cruse, Patricia, et al. *Web-Based Government Information: Evaluating Solutions for Capture, Curation, and Preservation*. Oakland, CA: California Digital Library, November 2003.

<http://www.cdlib.org/inside/projects/preservation/govinfo> (accessed 30 August 2004)

Dawes, Sharon. "The Role of Trust in New Models of Collaboration." In *New Models of Collaboration: A Guide for Managers*. Albany, NY: Center for Technology in Government, October 2003.

http://www.ctg.albany.edu/publications/online/new_models/essays/trust (accessed 30 August 2004)

Day, Michael. *Collecting And Preserving The World Wide Web: A Feasibility Study Undertaken For The JISC And Wellcome Trust*. Bath, UK: University of Bath, February 2003.

<http://library.wellcome.ac.uk/assets/WTL039229.pdf> (accessed 30 August 2004)

Decision Framework for Federal Document Repositories: Discussion Draft. Washington, DC: Prepared for the U.S. Government Printing Office by the Center for Research Libraries, April 12, 2004.

http://www.access.gpo.gov/su_docs/fdlp/pubs/decisionmatrix.pdf (accessed 30 August 2004)

DeMaio, Carl, et al. *Citizens Budget, 2003 – 2005*. Los Angeles, CA: Reason Public Policy Institute and the Performance Institute, 2003.

<http://www.rppi.org/citizensbudget.pdf> (accessed 30 August 2004)

Dickison, Meredith. *Persistent Locators for Federal Government Publications: Summary of a Study Conducted for the Depository Services Program and the National Library Of Canada*. Ottawa, ON: National Library of Canada, November 19, 2002.

<http://www.collectionscanada.ca/8/4/r4-500-e.html> (accessed 30 August 2004)

Dilevko, Juris and Elizabeth Dolan. *Government Documents Reference Service in Canada: Implications for Electronic Access*. Ottawa, ON: Communications Canada. Depository Service Program, March 20, 2002.

http://dsp-psd.communication.gc.ca/Rapports/Dilevko_Dolan/dilevko-e.html (accessed 30 August 2004)

Documents Data Miner ©: A Collection Management Tool For Federal Depository Libraries. Web site. Wichita, KS: Wichita State University.

<http://govdoc.wichita.edu/ddm/GdocFrames.asp> (accessed 30 August 2004)

Dolan, Elizabeth and Liwen Vaughan. *Electronic Access to Canadian Federal Government Information: How Prepared are the Depository Libraries?: Report to Depository Services Program, Canadian Government Publishing*. Ottawa, ON: Public Works and Government Services Canada, August 7, 1996.

<http://dsp-psd.communication.gc.ca/Rapports/DV/rep1-e.html> (accessed 30 August 2004)

Evans, T.C. "Current and Future State of GPO Access." Remarks at the Spring Federal Depository Library Council Meeting, April 22, 2002. Washington, DC: U.S. Government Printing Office, 2002.

http://www.access.gpo.gov/su_docs/fdlp/pubs/proceedings/02spc_te.pdf (accessed 30 August 2004)

The Fedora™ Project: An Open-Source Digital Repository Management System. Web site. Charlottesville, VA: University of Virginia, 2004.

<http://www.fedora.info> (accessed 30 August 2004)

Feldman, Susan and Chris Sherman. *The High Cost of Not Finding Information: An IDC White Paper*. Framington, MA: International Data Corporation (IDC), July 2001.

http://monkey.biz/Content/Default/Support/Resources/IDC_TheHighCostOfNotFindingInformation_1510.pdf (accessed 30 August 2004)

Final Report for The Pilot Project netarkivet.dk. Denmark: Royal Library, State Library, and the Centre for Internet Research, February 2003.

<http://www.netarkivet.dk/rap/index-en.htm>

Fisher, Janet. "Web Documents Digital Archive Pilot Project (OCLC)." In *Proceedings of the 10th Annual Federal Depository Library Conference*. Alexandria, VA, October 14 - 17, 2001.

http://www.access.gpo.gov/su_docs/fdlp/pubs/proceedings/01pro14.html (accessed 30 August 2004)

Fletcher, Patricia Diamond. "Leadership and Project Success: Lessons from High Impact Government Innovations." In *New Models of Collaboration: A Guide for Managers*. Albany NY: Center for Technology in Government, September 2003.

http://www.ctg.albany.edu/publications/online/new_models/essays/leadership (accessed 30 August 2004)

Gant, Dr. Jon P. "Delivering E-Government Services through the Access Indiana Information Network." In *New Models of Collaboration: A Guide for Managers*. Albany, NY: Center for Technology in Government, April 2003.

http://www.ctg.albany.edu/publications/online/new_models/cases/access_indiana.pdf (accessed 30 August 2004)

Georgia. *Official Code of Georgia Annotated. 2000 Supplement*, 20-5-2. Athens, GA: University of Georgia Libraries.

<http://www.libs.uga.edu/govdocs/collections/georgia/code.html> (accessed 30 August 2004)

Glasgow Caledonian University. *CASS Collaborative Academic Store for Scotland*. Web site. Glasgow, Scotland: Glasgow Caledonian University.

<http://www.lib.gcal.ac.uk/collections/cass.htm> (accessed 30 August 2004)

Gnassi, Bruno. *Accessing Canadian Federal Information: a Depository Program for the Twenty-First Century?*

Ottawa, ON: Communications Canada. Depository Service Program, March 2003.

<http://dsp-psd.communication.gc.ca/Rapports/IFLA/accessing-e.html> (accessed 30 August 2004)

Granger, Stewart, et al. *Cost Elements of Digital Preservation*. Version 4.0. Leeds, UK: University of Leeds, October 2000.

<http://www.leeds.ac.uk/cedars/colman/costElementsOfDP.doc> (accessed 30 August 2004)

Gupta, Amarnath. *Preserving Presidential Library Websites*. San Diego, CA: San Diego Supercomputer Center, January 18, 2001 (San Diego Supercomputer Center Technical Report)

<http://www.sdsc.edu/TR/TR-2001-03.pdf> (accessed 30 August 2004)

- Hartman, Cathy Nelson. *Collection Development Plan: Selecting Texas State Documents to be Included in the Electronic Depository Program*. Version 1.0. Austin, TX: Texas State Library, 2001.
<http://www.tsl.state.tx.us/lot/collectiondevelopmentplan.html> (accessed 30 August 2004)
- Hayes, Allene. "Web Archiving: New Developments." Presentation at the American Library Association Midwinter Conference, San Diego, California, January 11, 2004.
- Hedstrom, Margaret. *It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation: Final Report, Workshop on Research Challenges in Digital Archiving and Long-Term Preservation, April 12-13, 2002*. Washington, DC : Library of Congress, National Digital Information Infrastructure and Preservation Program, August 2003.
http://www.digitalpreservation.gov/repot/NSF_LC_Final_Report.pdf (accessed 30 August 2004)
- Horton, F. Woody. "Appendix 15. Some Important Information [on] Age Paradigms Shifts and Their Associated Myths and Realities." In *A Comprehensive Assessment of Public Information Dissemination*: Washington, DC: U.S. National Commission on Libraries and Information Science, 2001, vol. 3, p. 14-24.
<http://www.nclis.gov/govt/assess/assess.appen15.pdf> (accessed 30 August 2004)
- Illinois. *Illinois Compiled Statutes, State Library Act (15 ILCS 320/ from Ch. 128, par. 101)*. Springfield, IL: State of Illinois.
<http://www.cyberdriveillinois.com/library/isl/ref/readyref/addenda/citation.htm> (accessed 30 August 2004)
- Indiana University Bloomington Libraries. *Electronic Recordkeeping at Indiana University: NHPRC Funded Indiana University Electronic Records Project: Phase II, 2000-2002*. Bloomington, IN: Indiana University Bloomington Libraries, 2002
<http://www.indiana.edu/~libarch/ER> (accessed 30 August 2004)
- International Research on Permanent Authentic Records in Electronic Systems (InterPARES). *InterPARES 2 Project : International Research on Permanent Authentic Records in Electronic Systems*. Vancouver, BC: InterPARES Project, School of Library, Archival and Information Studies, The University of British Columbia, 2004.
<http://www.interpares.org> (accessed 30 August 2004)
- Ireland. Dept. of Enterprise, Trade and Employment. *BASIS Project - Content Management Study*. Dublin, Ireland: Ireland Department of Enterprise, Trade and Employment, June 2001.
<http://www.basis.ie/botaccess/doc.jsp?contentKey=WCCContent;id-31055&titleExtra=CM+Study> (accessed 30 August 2004)
- Jackson, Larry. *Preserving Electronic Publications: Final Report*. Springfield: Illinois State Library, 2002.
http://www.cyberdriveillinois.com/library/isl/lat/pep/pep_final.pdf (accessed 30 August 2004)
- Jacobs, James (Government Information Librarian, University of California, San Diego). E-mail to the March 2004 California State Library Government Documents Conference organizers, May 2004.
- James, Bruce R. *Keeping America Informed In The 21st Century: A First Look At The GPO Strategic Planning Process — "A Work In Progress."* Washington, DC: U.S. Government Printing Office, May 1, 2004.
http://www.access.gpo.gov/su_docs/fdlp/pubs/proceedings/James.DLC.04192004.revised.pdf (accessed 30 August 2004)
- Jones, Maggie. "Digital Preservation Activities in the U.K.: Building the Infrastructure." Paper presented at the World Library and Information Congress: 69th IFLA General Conference and Council, Berlin, Germany, 1-9 August 2003. The Hague: International Federation of Library Associations and Institutions, 2003.
<http://www.ifla.org/IV/ifla69/papers/129e-Jones.pdf> (accessed 30 August 2004)

Kadzabek, Susan, et al. *Staying Digital: Recommendations on Preserving New Jersey Government Information in the Digital Age*. Newark, NJ: State Documents Interest Group of the Documents Association of New Jersey, June 2001.

<http://newark.rutgers.edu/~govdocs/stayingdigital.pdf> (accessed 30 August 2004)

Kansas State Historical Society. *Kansas Electronic Recordkeeping Strategy: A White Paper*. Topeka, KS: Kansas State Historical Society, December 29, 1999.

<http://www.kshs.org/government/records/electronic/electronicrecordswhitepaper.htm> (accessed 30 August 2004)

"Kansas State Publications Archival Collection," *Crossroads: Developments in Electronic Records Management and Information Technology* 2003:2.

http://www.nagara.org/crossroads/2003_2.html (accessed 30 August 2004)

Kennedy, Linda and Rosita Lo Russo. "The Crisis in the California State Documents Distribution System." Unpublished paper, 1973.

Kenney, Anne, et al. "Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism," *D-Lib Magazine*, January 2002.

<http://www.dlib.org/dlib/january02/kenney/01kenney.html> (accessed 30 August 2004)

Koninklijke Bibliotheek (Netherlands). *The Archiving System For Electronic Publications: The E-Depot*. Web site. The Hague: National Library of The Netherlands.

http://www.kb.nl/kb/resources/frameset_kb.html?kb/dnp/e-depot/dm/dm.html (accessed 30 August 2004)

Koontz, Linda D. *Electronic Records: Management and Preservation Pose Challenges: Testimony Before the Subcommittee on Technology, Information Policy, Intergovernmental Relations, and the Census, Committee on Government Reform, House of Representatives*. Washington, DC: United States General Accounting Office, July 8, 2003.

<http://www.gao.gov/docdb/lite/summary.php?recflag=&accno=A07481&rptno=GAO-03-936T> (accessed 30 August 2004)

LaFon, Martin (Deputy Director, California Health and Human Services Data Center). Interview, 20 May 2004.

LC21: A Digital Strategy for the Library of Congress. Washington, DC: National Academy Press, 2000.

<http://www.nap.edu/books/0309071445/html> (accessed 30 August 2004)

Lamont, Melissa. "Here Today, Gone Tomorrow? Preserving Electronic Government Information for the Future." Paper presented at the 63rd IFLA General Conference, August 31- September 5, 1997, Copenhagen, Denmark. The Hague: International Federation of Library Associations and Institutions, 1997.

<http://www.ifla.org/IV/ifla63/63lamm.htm> (accessed 30 August 2004)

Larsen, Elena and Lee Rainie. *The rise of the e-citizen: how people use government agencies web sites*. Washington, DC: Pew Internet & American Life Project, 2002.

http://www.pewinternet.org/pdfs/PIP_Govt_Website_Rpt.pdf (accessed 30 August 2004)

Lavoie, Brian F. "Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century." *In Access in the Future Tense*. Washington, D.C.: Council on Library and Information Resources, 2004, p. 45-54.

<http://www.clir.org/pubs/reports/pub126/pub126.pdf> (accessed 30 August 2004)

- _____. *The Open Archival Information System Reference Model: Introductory Guide*. Dublin, OH: Online Computer Library Center, Inc. and Digital Preservation Coalition, 2004. (DPC Technology Watch Series Report; 04-01)
http://www.dpconline.org/docs/lavoie_OAIS.pdf (accessed 30 August 2004)
- _____. *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*. Dublin, OH: Online Computer Library Center, Office of Research, April 2003.
<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf> (accessed 30 August 2004)
- Library of Congress. *The MINERVA Web Preservation Project*. Web site. Washington, DC: Library of Congress.
<http://www.loc.gov/minerva> (accessed 30 August 2004)
- Library of Congress, et al., *September 11 Web Archive*. Web site. Washington, DC: Library of Congress MINERVA Project,
<http://september11.archive.org> (accessed 30 August 2004)
- LOCKSS for Government Documents: A Needs Assessment for the Federal Depository Library Community. Stanford, CA: Stanford University Libraries, January 2004.
<http://lockss-docs.stanford.edu/NAfinal.pdf> (accessed 30 August 2004)
- Lopresti, Robert. *Is All The Material Sent To Depository Libraries Available On The Web?* Issued in 4 parts. Bellingham, WA: Western Washington University Library, June 2000.
<http://www.library.wvu.edu/ref/subjects/govinfo/shipweb4.htm> [links directly to part 4; indirect links to parts 1-3] (accessed 30 August 2004)
- Lopresti, Robert and Marcia Gorin. "The availability of US government depository publications on the World Wide Web," *Journal of Government Information*. 29:1, pages 17-29 (January - February 2002).
- Lyman, Peter. "Archiving the World Wide Web." In *Building a National Strategy for Preservation Issues in Digital Media Archiving*. Washington, DC: National Digital Information Infrastructure and Preservation Program, Library of Congress, April 2002.
<http://www.clir.org/pubs/reports/pub106/web.html> (accessed 30 August 2004)
- Lynch, Clifford and Deanna Marcum. "Preservation of Scholarship: The Digital Dilemma." In *The Internet and the University: Forum 2002*. Boulder, CO: Educause, 2003, p. 21-24.
<http://www.educause.edu/ir/library/pdf/ffp0305s.pdf> (accessed 30 August 2004)
- Martin, Kristin and Jan Reagan. *North Carolina State Government Information: Realities and Possibilities*. Raleigh: State Library of North Carolina, November 2003.
<http://statelibrary.dcr.state.nc.us/digidocs/Workgroup/WhitePaper.pdf> (accessed 30 August 2004)
- Masanès, Julien. "Towards Continuous Web Archiving: First Results and an Agenda for the Future," *D-Lib Magazine*, 8:12 (December 2002).
<http://www.dlib.org/dlib/december02/masanés/12masanés.html> (accessed 30 August 2004)
- Matthews, Richard J., et al. *State-by-State Report on Permanent Public Access to Electronic Government Information*. Chicago, IL: American Association of Law Libraries, 2003.
http://www.ll.georgetown.edu/aallwash/State_PPAreport.htm (accessed 30 August 2004)
- Miszczynski, Dean (Director, California Research Bureau). Speech, California State Library Government Documents Conference, 23 March 2004.

National Archives of Australia. *Digital Records*. Web site. Canberra: National Archives of Australia. <http://www.naa.gov.au/recordkeeping/er/summary.html> (accessed 30 August 2004)

_____. *Managing Electronic Records: A Shared Responsibility*. Canberra: National Archives of Australia, 1997.
http://www.naa.gov.au/recordkeeping/er/manage_er/summary.html (accessed 30 August 2004)

National Association of Government Archives and Records Administration (NAGARA). Committee on Electronic Records and Information Systems (CERIS). *Status of the Preservation of Electronic Records by State Archives*. Albany, NY: National Association of Government Archives and Records Administrators, April 23, 2004.
http://www.nagara.org/news/ceris_report.pdf (accessed 30 August 2004)

National Library of Australia. *Digital Services Project*. Web site. Canberra: National Library of Australia. <http://www.nla.gov.au/dsp> (accessed 30 August 2004)

_____. *Keeping Government Publications Online: Guidelines for Commonwealth Publishers*. Canberra: National Library of Australia, July 2002.
<http://www.nla.gov.au/guidelines/govpubs.pdf> (accessed 30 August 2004)

_____. *PANDORA Archive: Preserving and Accessing Networked Documentary Resources of Australia*. Web site. Canberra: National Library of Australia and Partners.
<http://pandora.nla.gov.au/index.html> (accessed 30 August 2004)

National Information Standards Organization. *Data Dictionary – Technical Metadata for Digital Still Images* [draft standard]. Bethesda, MD: National Information Standards Organization, 2003.
http://www.niso.org/committees/committee_au.html (accessed 30 August 2004)

_____. *The Dublin Core Metadata Element Set*. Bethesda, MD: National Information Standards Organization, 2001.
<http://www.niso.org/standards/resources/Z39-85.pdf> (accessed 30 August 2004)

Nevada IT Operations Committee. e-Government Steering Committee. *Web Style Guide*. Carson City, NV: Nevada IT Operations Committee, 2002.
[http://www.nitoc.nevada.gov/NITOCdocs/3.02\(A\)_Policy_WebStyleGuide.pdf](http://www.nitoc.nevada.gov/NITOCdocs/3.02(A)_Policy_WebStyleGuide.pdf) (accessed 30 August 2004)

New Models of Collaboration: A Guide for Managers. Albany, NY: Center for Technology in Government, September 2003.
http://www.ctg.albany.edu/publications/online/new_models (accessed 30 August 2004)

New Zealand. E-Government Unit. *Metadata Management Facility User Requirements Specification: Portal Implementation Project*. Auckland, New Zealand: E-government Unit, State Services Commission, October 2001
<http://www.e-government.govt.nz/docs/mmf-users/index.html> (accessed 30 August 2004)

_____. *A Service Delivery Architecture for Government*. Auckland, New Zealand: E-Government Unit, State Services Commission, 2003.
<http://www.e-government.govt.nz/docs/service-arch-200303/index.html> (accessed 30 August 2004)

New Zealand Digital Library. Web site. Hamilton, New Zealand: University of Waikato, April 2000.
<http://www.cs.waikato.ac.nz/~nzdl/publications> (accessed 30 August 2004)

North Carolina. Division of Archives and History. *North Carolina Guidelines for Managing Public Records*. Raleigh, NC: North Carolina Department of Cultural Resources, Division of Archives & History, April 2000.

<http://www.ah.dcr.state.nc.us/e-records/manrecrd/manrecrd.htm> (accessed 30 August 2004)

_____. *Public Database Indexing: Guidelines and Recommendations, Release 1.1*. Raleigh, NC: North Carolina Department of Cultural Resources, Division of Archives and History, Records Services Branch, September 1996.

<http://www.ah.dcr.state.nc.us/e-records/pubdata/default.htm> (accessed 30 August 2004)

Pardo, Theresa A., et al. *Opening Gateways: A Practical Guide for Designing Electronic Records Access Programs*. Albany, NY: Center for Technology in Government University at Albany, SUNY. Second Edition, January 2002.

<http://www.ctg.albany.edu/resources/pdfrwp/gateways.pdf> (accessed 30 August 2004)

Pinfield, Stephen. "Open Archives and UK Institutions: An Overview." *D-Lib Magazine* 9:3 (March 2003).

<http://www.dlib.org/dlib/march03/pinfield/03pinfield.html> (accessed 30 August 2004)

Public Records Office Victoria (Australia). *Victorian Electronic Records Strategy (VERS Project) Final Report*. North Melbourne, Australia: Public Records Office Victoria, March 31, 1999

<http://www.prov.vic.gov.au/vers/pdf/final.pdf> (accessed 30 August 2004)

Rawan, Atifa. "Virtual Depository: Arizona Project Final Report and Recommendations." Presentation at the 12th Annual Federal Depository Library Conference, October 19-22, 2003, Arlington, Virginia. Washington, DC: U.S. Government Printing Office, 2003.

http://www.access.gpo.gov/su_docs/fdlp/pubs/proceedings/03pro_rawan.ppt (accessed 30 August 2004)

Reilly Jr., Bernard F. and Barbara DesRosiers. *Executive Summary, Developing Print Repositories: Models for Shared Preservation and Access*. Washington, DC: Council on Library and Information Resources, June 2003.

<http://www.clir.org/pubs/execsum/sum117.html> (accessed 30 August 2004)

Rudyard, Nigel. *Access For All: Online Access And Digitization*. London, UK: Commissioned by Resource, the Museums, Libraries and Archives Council for the Archives Task Force, September 2002.

http://www.mla.gov.uk/documents/atf_res01.pdf (accessed 30 August 2004)

Russell, Judy. "Future Directions of the Depository Library Program." Remarks made at the 142nd Association of Research Libraries Membership Meeting, May 15, 2003.

<http://www.arl.org/arl/proceedings/142/russell.html> (accessed 30 August 2004)

Russell, Judith C. *Remarks of Superintendent of Documents Judith C. Russell, Depository Library Conference St. Louis, Missouri*, April 18, 2004. Washington, DC: U.S. Government Printing Office, 2004.

http://www.gpoaccess.gov/about/speeches/04182004_DLC.pdf (accessed 30 August 2004)

San Diego Supercomputer Center. *The SDSC Storage Resource Broker*. Web site. San Diego, CA: San Diego Supercomputer Center.

<http://www.sdsc.edu/DICE/SRB> (accessed 30 August 2004)

Sanett, Shelby. "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks." *RLG DigiNews*, 7:4 (August 15, 2003).

<http://www.rlg.org/preserv/diginews/diginews7-4.html> (accessed 30 August 2004)

Searle, Sam and Thompson, Dave. "Preservation Metadata: Pragmatic First Steps at the National Library of New Zealand. " *D-Lib Magazine*, 9:4 (April 2003).

<http://www.dlib.org/dlib/april03/thompson/04thompson.html> (accessed 30 August 2004)

South Dakota. *South Dakota Statutes*, 2-16-8. Pierre, SD: South Dakota Legislature, Legislative Research Council.

<http://legis.state.sd.us/statutes/index.aspx?FuseAction=DisplayStatute&Type=Statute&Statute=2-16-8> (accessed 30 August 2004)

_____. *South Dakota Statutes*, 13-49-1. Pierre, SD: South Dakota Legislature, Legislative Research Council.

<http://legis.state.sd.us/statutes/index.aspx?FuseAction=DisplayStatute&Type=Statute&Statute=13-49-1> (accessed 30 August 2004)

Spooner, Tom. *Internet use by region in the United States*. Washington, DC: Pew Internet & American Life Project, 2003. http://www.pewinternet.org/pdfs/PIP_Regional_Report_Aug_2003.pdf (accessed 30 August 2004)

Sproull, Robert F. and Jon Eisenberg, eds. *Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for Initial Development*. Washington, DC: National Academies Press, 2003.

<http://www.nap.edu/books/0309089476/html> (accessed 30 August 2004)

Staples, Thornton, Ross Wayland and Sandra Payette. "The Fedora Project: An Open-source Digital Object Repository System," *D-Lib Magazine* 9:4 (April 2003).

<http://www.dlib.org/dlib/april03/staples/04staples.html> (accessed 30 August 2004)

State Library of Ohio. *Electronic Publications Pilot Project: Final Report*. Columbus, OH: State Library of Ohio, June 2001.

<http://www.ohiojunction.net/jerri/ep3final.pdf> (accessed 30 August 2004)

State Library of North Carolina. *Find NC: The Central Gateway to NC Government Information*. Web site. Raleigh, NC: State Library of North Carolina.

<http://www.findnc.org> (accessed 30 August 2004)

State Library of Tasmania. *Stable Tasmanian Open Repository Service*. Hobart, Tasmania: State Library of Tasmania, 2004.

<http://www.stors.tas.gov.au> (accessed 30 August 2004)

_____. *State Library of Tasmania: Our Digital Island*. Hobart, Tasmania : State Library, 2004.

<http://odi.statelibrary.tas.gov.au> (accessed 30 August 2004)

The State of Digital Preservation: An International Perspective; Conference Proceedings. Washington, D.C.: Council on Library and Information Resources, July 2002.

<http://www.clir.org/pubs/reports/pub107/contents.html> (accessed 30 August 2004)

Steenbakkens, Johan F. "Permanent Archiving of Electronic Publications: Research and Practice." In *International Summer School on the Digital Library 2003, Course 3: Libraries, Electronic Resources, and Electronic Publishing*. Tillburg, The Netherlands: Nationale Bibliotheek van Nederland, Koninklijke Bibliotheek, August 2003.

http://www.kb.nl/kb/hrd/dd/dd_links_en_publicaties/publicaties/summerschoolticer2003.pdf (accessed 30 August 2004)

- _____. "Treasuring the Digital Records of Science: Archiving E-Journals at the Koninklijke Bibliotheek." *RLG DigiNews*: Apr 15, 2004.
http://www.rlg.org/en/page.php?Page_ID=17068&Printable=1&Article_ID=990 (accessed 30 August 2004)
- Symons, George J. "Assessing the Value of Content: Information Management," *DM Review Magazine*, March 2004.
http://www.dmreview.com/article_sub.cfm?articleId=8175 (accessed 30 August 2004)
- Tuggle, Susan (Coordinator, Georgia Government Publications Program). Interview, 11 March 2004.
- Turk, Cathy and Adam Jensen. *Washington State Digital Archives Project Investment Plan*. Olympia, WA: Office of the Secretary of State, August 2003.
http://www.secstate.wa.gov/archives/pdf/digital_archives/InvestmentPlan.pdf (accessed 30 August 2004)
- U.S. Depository Library Council. *Depository Library Council's Advice to the Public Printer*. Washington, DC: Depository Library Council, Supt. Of Public Documents, US Government Printing Office, January 22, 2004.
http://www.access.gpo.gov/su_docs/fdlp/council/012204_council_rpt.pdf (accessed 30 August 2004)
- _____. *Depository Library Council, April 21, 2004*. Draft recommendations adopted at the Spring 2004 Depository Library Council Meeting, April 21, 2004, St. Louis, Missouri. Eugene, OR: Available from the Documents Interest Group of Oregon Site on the University of Oregon Libraries web page, 2004.
<http://libweb.uoregon.edu/govdocs/digor/DLC/2004-april-sum.html> (accessed 30 August 2004)
- U.S. Depository Library Council, Electronic Transition Committee. *Envisioning the Future of Federal Government Information : Summary of the Spring 2003 Meeting of the Depository Library Council to the Public Printer*. Washington, DC: U.S. Government Printing Office, 2003.
http://www.access.gpo.gov/su_docs/fdlp/council/EnvisioningtheFuture.html (accessed 30 August 2004)
- _____. *Report on GPO's Transition to a More Electronic FDLP*. Report at the Spring 2000 Depository Library Council Meeting. Newport, Rhode Island. Washington, DC: U.S. Government Printing Office, 2000.
http://www.access.gpo.gov/su_docs/fdlp/council/etreport.html (accessed 30 August 2004)
- U.S. Federal Interagency Committee on Government Information. Categorization of Government Information Working Group. *Defining What Government Information Is To Be Categorized: Statement of Requirements*. Washington, DC: U.S. Government Printing Office, May 13, 2004.
<http://www.gpoaccess.gov/cgiwg/pdf/cgiwgroup/revMay2004.pdf> (accessed 30 August 2004)
- U.S. Government Printing Office. *Collection Of Last Resort: Discussion Draft, April 6, 2004*. Washington, DC: U.S. Government Printing Office, April 6, 2004.
http://www.access.gpo.gov/su_docs/fdlp/pubs/clr.pdf (accessed 30 August 2004)
- _____. *Report on the Meeting of Experts on Digital Preservation*. Washington, DC: U.S. Government Printing Office, March 12, 2004.
<http://www.gpoaccess.gov/about/reports/preservation.pdf> (accessed 30 August 2004)
- U.S. Government Printing Office. Library Programs Service. *Managing the FDLP Electronic Collection: A Policy and Planning Document*. Washington, DC: Library Programs Service, Supt. of Documents. US Government Printing Office, October 1, 1998.
http://www.access.gpo.gov/su_docs/fdlp/pubs/ecplan.html (accessed 30 August 2004)
- U.S. National Commission on Libraries and Information Science. "Appendix 25, Panel 3. Final Report on Citizen, Business, Lower Levels of Government, Library, and Other Needs for Public Information

Products and Services.” In *A Comprehensive Assessment of Public Information Dissemination*: Washington, DC: U.S. National Commission on Libraries and Information Science, 2001, vol. 3, p. 124-135.

<http://www.nclis.gov/govt/assess/assess.appen25.pdf>(accessed 30 August 2004)

_____. *Report On The Assessment Of Electronic Government Information Products*. Washington, DC: U.S. Government Printing Office, March 30, 1999.

http://www.access.gpo.gov/su_docs/nclisassessment/report.html (accessed 30 August 2004)

United Kingdom. Cabinet Office. E-Government Unit. *e-Government Unit Guidelines*. London, UK: Cabinet Office, e-Government Unit, 2004.

<http://www.e-envoy.gov.uk/Resources/Guidelines/fs/en> (accessed 30 August 2004)

United Kingdom. National Archives. *Preservation: Digital Archives*. Web site. Kew, Richmond, Surrey (UK): National Archives, Public Record Office.

<http://www.pro.gov.uk/about/preservation/digital/archive/default.htm> (accessed 30 August 2004)

United Kingdom. Prime Minister. *Modernizing Government: Presented to Parliament by the Prime Minister and the Minister for the Cabinet Office by Command of Her Majesty*. London : The Stationery Office, March 1999.

<http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm> (accessed 30 August 2004)

United Kingdom. Stationery Office. *Legal Deposit Libraries Act: 2003 Chapter 28*. London: Her Majesty's Stationery Office, November 3, 2003.

<http://www.legislation.hms.gov.uk/acts/acts2003/20030028.htm> (accessed 30 August 2004)

_____. *Official Documents Website*. London, UK: The Stationery Office.

<http://www.official-documents.co.uk/index.html> (accessed 30 August 2004)

United States. U.S. Code, *17 USCS Sects. 401 – 412* [copyright law]. Ithaca, NY: Cornell Law School, Legal Information Institute.

<http://www.law.cornell.edu/copyright/17usc407> (accessed 30 August 2004)

_____. *U.S. Code, Title 44, Public Printing and Documents, Chapter 31, Records Management by Federal Agencies*. Washington, DC: U.S. Government Printing Office.

http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1994_uscode_suppl_3&docid=44usc3101 (accessed 30 August 2004)

University of Arizona Library. *Government Printing Office Pilot Project (Electronic Depository)*. Tucson, AZ: University of Arizona Library, September 2002.

<http://www.library.arizona.edu/library/teams/sst/pol/guide/gpo-pilot/index.html> (accessed 30 August 2004)

University of Georgia. Government Documents Department. *Depository Requirements: State of Georgia Publications*. Athens, GA: University of Georgia Libraries, Government Documents Department.

<http://www.libs.uga.edu/govdocs/collections/georgia/require.html> (accessed 30 August 2004)

University of Glasgow. *Effective Records Management (ERM) Project*. Web site. Scotland: University of Glasgow.

<http://www.gla.ac.uk/InfoStrat/ERM> (accessed 30 August 2004)

University of Illinois at Urbana-Champaign. Information Systems Research Lab. *Preserving Electronic Publications (PEP)*. Web site. Champaign: University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science.

<http://www.isrl.uiuc.edu/pep> (accessed 30 August 2004)

University of North Texas Libraries. *CyberCemetery*. Web site. Denton, TX: University of North Texas.
<http://govinfo.library.unt.edu/> (accessed 30 August 2004)

Utah State Archives. *Electronic Records Page*. Web site. Salt Lake City: Utah State Archives.
<http://www.archives.state.ut.us/recmanag/electronic.htm> (accessed 30 August 2004)

Van Dijk, Marcel. "It Always Hurts The First Time: Experiences With Transferred Electronic Records,"
Cultivate Interactive, Issue 9 (February 2003).
<http://www.cultivate-int.org/issue9/amsterdammro> (accessed 30 August 2004)

Warner, Dorothy. "'Why Do We Need to Keep This In Print? It's on the Web:' A Review of Electronic Archiving Issues and Problems," *Progressive Librarian*, no. 19-20, Spring 2002.
http://libr.org/PL/19-20_Warner.html (accessed 30 August 2004)

Washington (State). *Revised Code of Washington, RCW 40.14.020*. Olympia, WA: Washington State Legislature.
<http://www.leg.wa.gov/RCW/index.cfm?section=40.14.020&fuseaction=section> (accessed 30 August 2004)

Witten, Ian H. "Examples of Practical Digital Libraries Collections Built Internationally Using Greenstone,"
D-lib Magazine, 9:3 (March 2003).
<http://www.dlib.org/dlib/march03/witten/03witten.html> (accessed 30 August 2004)

The Xtensible Past XML as a means for easy access to historical research data and a strategy for digital preservation (NIWI). Amsterdam: Netherlands Institute for Scientific Information Services, September 2003.
http://www.niwi.knaw.nl/nl/geschiedenis/projecten/xpast_copy1/toon (accessed 30 August 2004)

Appendix 4

List of Interviews

APPENDIX 4

Managing and Sustaining A State Government Documents Program in California

Interviews conducted for the report

ARIZONA

Richard Pearce-Moses, Director, Digital Government Information, Arizona State Library, Archives and Public Records. Interview, March 8, 2004.

Tom Martin, Chief Information Officer, Arizona State Library, Archives and Public Records. Interview, March 8, 2004.

CALIFORNIA

Lucy Barber, Archivist and Walter Gray, State Archivist, California State Archives. Interview, December 4, 2003.

Guy Blair, Administrative Manager, Office of State Publishing and staff members, Teresa Squier, Mark Renteria. Interview, December 4, 2003.

Heather Christenson, California Digital Library. Interview, January 10, 2004.

Janet Coles, Senior Librarian, Government Publications Section, California State Library. Written interview questions, November 24, 2003; January 21, 2004; February 2, 2004.

Stephanie Couch, Director, DCP Statewide Initiatives, CENIC/Digital California Project. Telephone interview, May 19, 2004.

Patricia Cruse, Director, Digital Preservation Program, California Digital Library. Interview, January 10, 2004; telephone interview, May 13, 2004.

David Harris, Internet Services Coordinator, CERES (California Environmental Resources Evaluation System). Interview, May 11, 2004.

John Jewell, Chief, State Library Services Bureau, California State Library. Telephone interview, February 19, 2004

Martin LaFon, Deputy Director, California Health and Human Services Data Center. Telephone interview, May 20, 2004.

Terry Roberts, Director, State Clearinghouse and Planning Unit, California Governor's Office of Planning and Research. Telephone interview, May 18, 2004.

Group interview, Debbie Newton, Dennis Hagen, Jackie Siegel, Information Technology Bureau, California State Library. December 4, 2003.

Group meeting with Linda Kennedy (University of California, Davis), Charles Eckman (Stanford University), and other University of California and California State University documents librarians. January 10, 2004.

GEORGIA

Susan Tuggle, Coordinator of Georgia Government Publications, University of Georgia Libraries. Telephone interview, March 11, 2004.

ILLINOIS

Larry Strickland, University of Illinois, Urbana/Champaign. Interview, March 8, 2004.

NORTH CAROLINA

Kristin Martin, Digital State Documents Librarian, State Library of North Carolina. Interview, March 8 2004.

U.S. GOVERNMENT PRINTING OFFICE

George Barnum, Electronic Collection Manager, U.S. Government Printing Office. Interview, April 13, 2004.

WASHINGTON

Adam Jansen, Digital Archivist, State of Washington, Office of the Secretary of State. Phone interview, March 18, 2004.

Appendix 5—List of Acronyms

CalRIM	California Records and Information Management
CDL	California Digital Library
CENIC	Corporation for Education Network Initiatives in California
CERES	California Environmental Resources Evaluation System
CPR	California Performance Review
CSL	California State Library
DCMS	OCLC Digital Collections and Metadata Services Division
GPO	U.S. Government Printing Office
GPS	Government Publications Section (California State Library)
LDA	Library Distribution Act
LOCKSS	Lots of Copies Keep Stuff Safe
OAIS	Open Archival Information System
OSP	California Office of State Publishing
NARA	U.S. National Archives and Records Administration
NISO	National Information Standards Organization
SAM	State Administrative Manual
SDSC	San Diego Supercomputer Center
US GPO	U.S. Government Printing Office